



**Universidade de
Aveiro
2005**

Departamento de Línguas e Culturas

**Silvana Marta Pinho
Paiva**

**Síntese por concatenação de variantes regionais:
falar do Porto**



**Universidade de
Aveiro
2005**

Departamento de Línguas e Culturas

**Silvana Marta Pinho
Paiva**

**Síntese por concatenação de variantes regionais:
falar do Porto**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Estudos Portugueses, realizada sob a orientação científica da Doutora Lurdes de Castro Moutinho, Professora Associada do Departamento de Línguas e Culturas e do Doutor António Teixeira, Professor Auxiliar do Departamento de Electrónica e Telecomunicações da Universidade de Aveiro.

O júri

Presidente

Prof. Doutor Francisco António Cardoso Vaz
Professor Catedrático da Universidade de Aveiro

Vogais

Profa. Doutora Lurdes de Castro Moutinho
Professora Associada da Universidade de Aveiro (orientadora)

Prof. Doutor João Manuel Pires da Silva e Almeida Veloso
Professor Auxiliar da Faculdade de Letras da Universidade do Porto

Profa. Doutora Rosa Lúcia Torres de Couto Coimbra e Silva
Professora Auxiliar da Universidade de Aveiro

Prof. Doutor António Joaquim da Silva Teixeira
Professor Auxiliar da Universidade de Aveiro (Co-orientador)

Agradecimentos

Gostaria de expressar a minha gratidão a todos quantos, por diferentes formas, me ajudaram na elaboração deste trabalho.

Dirijo o meu sincero reconhecimento e agradecimento aos Professores Doutores Lurdes de Castro Moutinho e António Teixeira pela orientação, apoio, disponibilidade, compreensão, incentivo, profissionalismo e as suas palavras amigas que em todas as circunstâncias me proporcionaram.

Palavras – chave

Fonética experimental, variação dialectal, a variante do Porto, difones, síntese por concatenação.

Resumo

O desenvolvimento de uma nova voz, característica da variante do Porto, utilizando a síntese por concatenação de difones é apresentado nesta dissertação.

Na primeira parte, descrevemos os estudos realizados no último século, sobre a unidade e diversidade dialectal da Língua Portuguesa, onde nos referimos apenas aos dialectos continentais – Setentrional e Centro-Meridional.

A variedade escolhida para esta é o falar do Porto, o qual se integra no dialecto Setentrional. Esta variedade, bem como os seus fenómenos fonéticos característicos, serviram-nos de base para a construção da lista de difones a contemplar na criação desta voz.

A segunda parte, consiste na construção da voz do Porto, usando as ferramentas e os módulos disponibilizados pelo sistema Festival.

Finalmente, avaliamos a voz através de testes perceptuais, tentando validar a sua inteligibilidade e a sua aceitação por parte dos ouvintes.

Keywords

Experimental Phonetics, dialectal variation, Porto dialect, diphones, concatenative synthesis.

Abstract

A Porto dialect concatenation synthesis system is presented as a case study in prototyping TTS for resource-poor minority dialects. For these varieties, demonstrator systems are essential to seed projects in speech and language technology.

In the first part, a description of the studies realized during last century about Continental Dialects is presented. It follows some information about the main phonetic characteristics of these dialect, nowadays.

The last part consist in the Porto Voice building according to the tools and modules of Festival system.

An evaluation, using perceptual tests, with speakers acknowledging Porto dialect shows that such a hybrid prototype can already produce intelligible speech for demonstration purposes.

Índice

1	Introdução	1
1.1	Objectivos	2
1.2	Resultados Publicados	2
1.3	Estrutura.....	3
2	Variação Dialectal	5
2.1	- A Unidade e Diversidade do Português Europeu.....	5
2.1.1	Dialectos Centro-Meridionais.....	9
2.1.2	Dialecto Setentrional	9
2.2	Um caso particular dos Dialectos Setentrionais: a variante do Porto	10
2.2.1	Alguns estudos sobre o falar do Porto	11
2.3	O Falar do Porto: breve estudo sobre a situação actual.....	13
2.3.1	Traços diferenciadores	14
3	Sobre Síntese e Sistemas de Síntese de Voz.....	21
3.1.1	Síntese: o que é?	21
3.2	Estrutura dos sistemas TTS	22
3.3	Um pouco de história	22
3.4	Aplicações.....	24
3.5	Tipos de síntese.....	25
3.6	Síntese por concatenação	26
3.7	Síntese do Português	27
3.8	O sistema Festival.....	28
4	Construção de uma voz do Porto para o sistema <i>Festival</i>	31
4.1	Descrição Geral	31
4.2	A escolha da lista de fones.....	32
4.3	Construção da lista de palavras contendo os difones.....	33
4.4	Gravação do corpus	35
4.5	Anotação do corpus	36
4.6	Construção do índice de difones	38
4.7	Extracção da informação dos períodos glotais e parâmetros LPC	38
4.8	Duração dos fones.....	39
4.9	Teste e verificação da base de dados	39
5	Resultados e discussão.....	41
5.1	A avaliação das vozes	41
5.1.1	Procedimento	41
5.1.2	Estímulos	43
5.1.3	Ouvintes	43
5.1.4	Resultados para a voz com pronúncia standard	44
5.1.5	Discussão	58
5.2	Teste perceptual para avaliação da qualidade da voz.....	59
5.2.1-	Procedimento	59
5.2.2-	Estímulos	60
5.2.3-	Os ouvintes	62
5.2.4-	Os resultados	64
6	Conclusões	77
6.1	Resumo do trabalho efectuado.....	77
6.2	Resultados principais	78

6.3 Trabalho Futuro	79
Bibliografia	81
Endereços electrónicos	84
Anexo 1	85
Anexo 2	88
Anexo 3	94
Anexo 4 - frases tratadas para o teste 6.2	100

1 Introdução

Com o advento do processamento computacional da Língua Portuguesa, onde se inclui a síntese de fala, torna-se necessário criar condições para que, também neste âmbito, as variantes sejam consideradas. Um dos pontos-chave na aceitação dos sistemas de voz pela generalidade dos utilizadores, condição para a utilização generalizada desta tecnologia, prende-se com a necessidade dos sistemas estarem o mais próximo possível dos utilizadores, isto é, serem capazes de reproduzir as variantes próprias de um determinado espaço dialectal.

Concretamente, é nosso objectivo criar uma voz, de acordo com uma variante dialectal, a do Porto, para o sistema Festival, sistema aberto e com facilidade de adição de vozes para novas línguas, usando a síntese por concatenação de difones, uma das técnicas mais habituais na área da síntese.

A criação desta voz servirá como instrumento de trabalho na área da fonética experimental, na medida em que pode contribuir para um aprofundamento do estudo dos fenómenos linguísticos específicos da região em estudo. Para além disso, pode também servir como base de dados para posteriores estudos sobre essa mesma variedade, bem como vir a ser utilizada, como ponto de partida para a criação de “novas” vozes, de acordo com outras variantes e/ou variedades.

O trabalho confronta alguns aspectos conhecidos da variedade normativa do Português Europeu (PE) com a variedade do Porto. Desta forma, foram seleccionados dois informantes: um, com realizações fonéticas características do falar do Porto; um outro que, não possuindo nenhuma marca própria de um qualquer falar, foi por nós considerado um representante da variedade padrão.

Uma vez que este trabalho descreve e analisa fenómenos linguísticos que estão sujeitos a grande variação (dos quais os falantes podem ou não ter consciência), não pode, nem pretende ser, um estudo exaustivo da variedade dialectal do Porto. O estudo procura, sobretudo, salientar a importância que a variação linguística pode ter, nos estudos linguísticos e nos estudos de síntese de voz.

1.1 Objectivos

Ao iniciarmos este trabalho torna-se pertinente enunciar os objectivos que nos propomos atingir, sendo o principal, e tema central desta dissertação, a criação de uma voz sintética característica de um falar do Português Europeu, neste caso o falar do Porto. Pretende-se desta forma contribuir, mesmo que modestamente, para que haja cada vez mais vozes sintéticas relativas à Língua Portuguesa. Só com a existência de uma maior oferta poderá existir uma utilização em massa deste tipo de tecnologia.

Esta dissertação pretende averiguar se, mesmo no desenvolvimento de sistemas de processamento automático da Língua Portuguesa, neste caso a síntese, a variação regional é um factor a ter em consideração. Para isso, propusemo-nos confrontar um sistema baseado numa voz standard, cuja pronúncia foi adaptada, mantendo-se uma voz standard, com um sistema integrando uma voz desenvolvida tendo em conta as características do falar do Porto, aproveitando o estudo para aferir da validade actual das descrições dessa variante.

É também nossa finalidade disponibilizar uma ferramenta que permita a realização de experiências relativas, por exemplo, à produção de certas palavras em que exista alguma discrepância de realização e transcrições publicadas.

O desenvolvimento da voz, surge, assim, não só como objectivo em si, mas também como a ferramenta, ou mesmo, a motivação para a realização de outros objectivos.

1.2 Resultados Publicados

Resultados parciais obtidos no decorrer desta pesquisa foram já apresentados no XX Encontro da Associação Portuguesa de Linguística e serão publicados nas actas (Paiva, et al., no prelo).

A co-autoria do trabalho *‘Um novo sistema de conversão grafema-fone para o PE baseado em transdutores’*, apresentado no II Congresso Internacional de Fonética e Fonologia em 2004, em S. Luís do Maranhão (Oliveira, C. et al., no prelo), decorre, também, do trabalho inicial, relativo às regras de conversão grafema-fone para o falar do Porto.

1.3 Estrutura

Para levarmos a cabo esta investigação, foram realizadas diferentes tarefas que constituem os diferentes capítulos da dissertação. De forma a estruturar a informação, a progressão da pesquisa e os resultados obtidos, o nosso trabalho será, então, dividido em capítulos, organizados como seguidamente se apresenta:

O capítulo 1 inclui uma introdução, apresentando, de seguida, os objectivos e a estrutura desta dissertação.

No segundo capítulo, intitulado Variação Dialectal, faremos referência à unidade e diversidade do Português Europeu, incluindo as seguintes secções:

- * Dialecto Centro-Meridional
- * Dialecto Setentrional
- * Um caso particular dos Dialectos Setentrionais: variante do Porto

Os dialectos Centro-Meridional e Setentrional serão referidos de uma forma muito sucinta, uma vez que não são o objecto do nosso estudo.

Contudo, como a variante do Porto surge, inicialmente, como um caso particular do Dialecto Setentrional, faremos alusão a alguns estudos realizados sobre esse falar. Os traços diferenciadores do Falar do Porto serão objecto de estudo e análise. Esses traços, sendo a base da nossa dissertação, serão de seguida avaliados/validados com um estudo de caso. Este não só nos permitirá validar a existência destes fenómenos, mas também, testar a sua actualidade.

Questões relacionadas com síntese e sistemas de síntese de voz constituirão o terceiro capítulo. Inicialmente, faremos uma breve alusão à sua história, desde o seu aparecimento até aos sistemas recentes, seguindo-se uma descrição sumária dos tipos de síntese, da síntese por concatenação e do sistema Festival.

O capítulo 4 será consagrado à construção de uma voz do Porto para o sistema Festival. Neste capítulo, apresentaremos uma descrição geral de todo o processo, seguida dos vários passos necessários à construção da voz.

Os resultados de dois testes perceptuais, para avaliação da voz criada no que diz respeito à qualidade e à percepção (como sendo do falar do Porto), serão apresentados e discutidos no quinto capítulo.

No capítulo 6, que intitulámos de Conclusões, teceremos alguns comentários finais à pesquisa efectuada, referindo os principais resultados e apontando para pesquisas futuras.

Este último capítulo será seguido da bibliografia por nós consultada, bem como de alguns endereços electrónicos relevantes.

Finalmente, em forma de anexos, serão apresentadas todas as tabelas e outros documentos que consideramos pertinentes para a leitura e análise desta dissertação.

2 Variação Dialectal

2.1 - A Unidade e Diversidade do Português Europeu

O Português é uma língua consensualmente caracterizada por uma grande unidade das suas variedades dialectais, sobretudo as peninsulares. Essa unidade é visível no facto de todos os falantes se entenderem sempre, qualquer que seja a sua origem.

Apesar da proximidade reconhecida entre as variantes do Português Europeu (PE), existem características sentidas pelos falantes de uma determinada proveniência como típicas de outras variantes. Paralelamente, existem fenómenos variáveis em cada variante, e mesmo no falar de cada indivíduo, que carecem de integração no âmbito das descrições do conhecimento linguístico.

O PE não tem sido objecto de muitos estudos centrados na variação, ao contrário do que acontece com o Português do Brasil (PB) e com outras línguas como o Inglês, por exemplo. Na verdade, são poucos os estudos actuais neste domínio e, por essa razão, entendemos que deveríamos debruçar-nos sobre o assunto.

No caso do Português Europeu, apesar de ser considerada uma Língua bastante homogénea, quando comparada a outras Línguas, igualmente românicas, a existência de fenómenos de variação linguística, mais facilmente detectáveis em termos fonéticos, é notória, tanto no território continental, como insular. Apesar disso, de entre as variedades consideradas, existe uma que, por motivos puramente extra-linguísticos (de ordem política, histórica, entre outros) se impôs, como a Variedade Normativa. Será sempre como referência a esta última que analisamos os fenómenos específicos de cada variedade e a forma como cada uma delas diverge da variante considerada prestigiante.

Essa variedade normativa é sempre definida como um «dialecto de prestígio», característico de um grupo social e de um centro cultural considerado modelar, não por razões estritamente linguísticas, mas por razões sócio-políticas e culturais. Além dessa variedade de prestígio, resultado, em geral, do consenso entre letrados, académicos,

gramáticos, existe em toda comunidade de fala, de entre todas as suas variações típicas, uma norma consensual que não é imposta de fora ou de cima, mas resultado da necessidade de intercomunicação interdialectal, criando-se, assim, uma espécie de unidade dentro da diversidade linguística.

Estes fenómenos de variação foram encarados até aos anos 50/60 como marginais, sendo mesmo considerados desprestigiantes. Com o advento da sociolinguística, nomeadamente da escola americana de Labov, entre outras, ao trabalhar sistematicamente o aspecto social da língua, a variação linguística passa a ser encarada sob uma outra perspectiva. Este tipo de pesquisas, ao trazerem a língua para o seu contexto social em situações de uso, fazem com que, concomitantemente, surja também uma maior aceitação da diversidade linguística. A variedade normativa, até aqui considerada “a única língua”, passa a ocupar o lugar entre outras variedades mantendo-se, apesar disso, como referência, mas não excluindo as outras variedades existentes.

Um «dialecto de prestígio» sobre o qual se erige a norma oficial, em termos estritamente linguísticos, tem a mesma qualificação de qualquer outro dialecto de uma língua histórica. O seu carácter de prestígio advém, portanto, de factores extralinguísticos. Na linguística contemporânea, o termo dialecto não mais carrega a qualificação estigmatizada, comum a fases passadas da história do estudo sobre a linguagem. A «língua de cultura» era, definida então como a que dava suporte à manifestação escrita, transmitida sobretudo pelos processos de escolarização, de aculturação, portanto, em direcção ao grupo social considerado modelo sócio-cultural da sociedade.

Por dialecto, entende-se hoje, as variedades de uma língua histórica que caracterizam formas de falar específicas de lugares, estratos sociais, faixas etárias, com seus registos próprios e adequados à formalidade da situação de comunicação.

O dialecto chamado de prestígio ou standard não é mais do que um dos «modos de falar» (significado etimológico de dialecto, do grego «*dialektos*») de uma determinada língua histórica, próprio a um grupo social de um determinado lugar, que é, pela sociedade em que se insere, escolhido como modelo a ser seguido. Esse modelo é obrigatório, seguido em certas instâncias e, neste caso, poderia ser visto como uma das múltiplas formas de controlo social existentes nas sociedades humanas, com o fim político de neutralizar a diversidade natural às sociedades e línguas históricas.

Apesar da aparente homogeneidade da Língua Portuguesa desde há longas décadas que é reconhecida a existência de variação PE, pelo que, dialectólogos portugueses, com base nas suas investigações, constituíram o mapa dialectológico dos dialectos continentais. Nesses mapas, são apresentadas as variantes encontradas em cada uma das áreas linguísticas por eles definidas.

Uma vez que vamos tratar de variedades ou falares, é importante distinguirmos língua, dialecto e falar.

Paiva Boléo e Maria Helena Santos Silva (1959 (1962)), no Mapa dos Dialectos e Falares de Portugal, fazem já a distinção entre dialectos, falares e variedades. Boléo reconhece os dialectos mirandês, rioronês e guadranês, considerando todos os outros como falares. Leite Vasconcelos (1901) refere o dialecto interamnense, o dialecto transmontano, o dialecto beirão e o dialecto meridional.

A distinção feita por Celso Cunha e Lindley Cintra (1984), em tudo se assemelha a de Manuel Alvar (1961), ao afirmar que «as formas características que uma língua assume regionalmente denominam-se dialectos». Para Cintra e Cunha (1984) «falar seria a peculiaridade expressiva própria de uma região e que não apresenta o grau de coerência alcançado pelo dialecto. Assim, tudo o que não era padrão teria uma denominação própria, sendo também, e por isso, um pouco negligenciado. Isto é, não pode ser aceite, porque é diferente da norma, algo inferior e secundário.

De acordo com os estudos dialectológicos realizados por Cintra (1971(1983)), podem ser identificados em Portugal Continental dois grupos de Dialectos, com base em diferenças fonéticas bem definidas: o “português setentrional e português centro-meridional” (op.cit., p.145). Ainda que, em cada dialecto, o autor reconheça subdivisões, a verdade é que toda a zona, numa linha transversal, passando a Norte de Aveiro e abaixo de Castelo Branco pertenceria ao dialecto Setentrional. O restante território continental estava inserido no dialecto Centro-meridional.

O mapa dialectológico proposto por Cintra é apresentado na figura 1.



Figura 1- Os dialectos portugueses segundo Luís Filipe Lindley Cintra (mapa adaptado de Cintra 1971 por Segura e Saramago 2001)

Apesar da existência de algumas diferenças quanto ao nome atribuído aos dialectos, ao longo do tempo, parece-nos, actualmente, existir um consenso na aceitação da classificação proposta por Lindley Cintra (1971(1983)), pelo que a adoptámos no presente estudo.

De acordo com esta escolha, passamos, de seguida, a apresentar, de forma sumária, os dialectos Centro Meridional e Setentrional, bem como algumas das suas características.

Após uma breve caracterização do dialecto Setentrional, abordaremos o Falar do Porto, como um caso particular deste dialecto e que, por isso, tem merecido, nas classificações dialectológicas, um lugar de destaque. Numa primeira abordagem, citaremos alguns estudos realizados sobre este Falar, referindo alguns dos seus fenómenos fonéticos mais característicos. Seguidamente, apresentaremos um breve estudo de um caso exemplar,

por nós realizado, através do qual é possível mostrar a actualidade das realizações fonéticas características do Falar do Porto, confirmando-se, deste modo, a pertinência deste estudo.

2.1.1 Dialectos Centro-Meridionais

Nos Dialectos Centro-Meridionais encontramos, do ponto de vista dialectal, uma maior homogeneidade. Podemos, contudo, fazer a distinção entre os dialectos do Centro-Litoral (estremenho e beirões) e os dialectos do Centro-Interior (ribatejano, baixo-beirão, alentejanos e algarvios). A fronteira que os separa é a isófona que corresponde à monotongação do ditongo [ej] na vogal [e]. Esta fronteira parte da costa ocidental de Óbidos, desce até quase à zona de Lisboa e, subindo posteriormente para Nordeste, acompanha o curso do rio Zêzere, atingindo a área norte de Castelo Branco.

É de salientar a especificidade da zona de Lisboa que constitui um centro de excepção nesta grande área de monotongação de [ej] para [e], optando por manter o ditongo. Mais a norte, este ditongo é realizado como [6j], como afirma Cintra (1983, p.29-30) “Todo o Norte e grande parte do centro conservam o ditongo aumentando por vezes a distância entre os seus elementos, por dissimilação: ferrâiro, lâite. Está neste caso a cidade de Lisboa, ilhota de conservação de [ej] numa zona em que o ditongo desapareceu”.

2.1.2 Dialecto Setentrional

Uma das características que distingue o Dialecto Setentrional é a manutenção do ditongo [ow], que a norma monotongou em [o]. Palavras como *pouco*, *louco* e *touro* são realizadas [‘powku], [‘lowku] e [‘towru], em vez das normalizadas [‘poku], [‘loku] e [‘toru]. O ditongo [ow] no Dialecto Centro - Meridional é realizado sempre monotongado.

No Baixo-Minho e Douro Litoral, ocorrem dois traços particulares que se prendem com: 1) “a ditongação das vogais médias acentuadas [e] e [o] respectivamente em [je] e [wo] ou mesmo [w6] ¹, como em [‘pjezu] por [pezu] (*peso*) ou [s@’Jwor6], [s@’Jw6r6] por [s@Jor6] (*senhora*); 2) a conservação da terminação – om, que a norma substitui por –

¹ Nas diferentes transcrições apresentadas, utilizaremos o sistema de transcrição fonética SAMPA para o Português, cujos símbolos serão apresentados em anexo(anexo 1).

ão, bem como a sua generalização a todas as palavras terminadas em –ão” (Faria 1996, p. 495), como em *pão* [p6~w~] e *cão* [k6~w], que se realizam [po~] e [ko~].

2.2 *Um caso particular dos Dialectos Setentrionais: a variante do Porto*

Nos dialectos Setentrionais, destaca-se o falar do Porto, que é sempre merecedor de uma atenção particular nos vários estudos dialectais. No entanto, as suas características específicas não têm sido exaustivamente estudadas, muito menos merecido estudos experimentais. Apesar desses estudos lacunares, os traços fonéticos que caracterizam este falar são comumente aceites e intuitivamente referidos por portuenses e falantes de outras variedades.

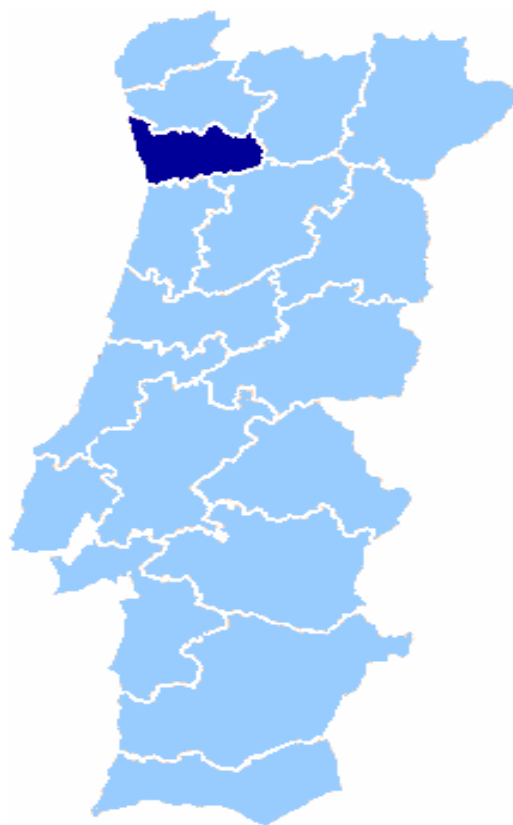


Figura 2- Mapa de Portugal Continental, com destaque para o distrito do Porto (assinalado com um tom mais escuro)

Para além destes aspectos particulares de linguagem, o Porto, como património nacional e mundial, é, por si só, motivo de outros estudos e abordagens científicas.

De forma muito sucinta, descreveremos um pouco da história desta cidade, não querendo, no entanto, fazer deste tema o objecto da nossa tese. Consideramos, contudo, pertinente abordá-lo, pois, apesar de toda a evolução industrial e tecnológica e das suas facetas de modernidade, o Porto mostra-se relutante em abdicar de aspectos mais conservadores, o que parece também verificar-se com o seu próprio falar.

Na verdade, ao longo do tempo, o Porto conheceu mudanças urbanas e sociais que, a partir do séc. XVIII, sob impulso do comércio do vinho e ao longo do séc. XIX, com a industrialização, operaram radicais alterações na sua fisionomia e na maneira de ser dos habitantes. Tais mudanças viriam a acentuar-se, no século passado, particularmente nas últimas décadas. Todavia, apegada a um instinto vital de sobrevivência, a cidade guardou o culto e a admiração pelos testemunhos da particularidade do seu génio, das realizações artísticas do seu passado, da autonomia do seu espírito inventivo, das suas tradições. Inovadora e conservadora ao mesmo tempo, esta cidade orgulha-se do seu património cultural. A sua riqueza e beleza arquitectónica, os seus espaços culturais onde todos os dias surgem novos incentivos à sua cultura e passado históricos, bem como os seus espaços de lazer transformaram o Porto em Património Mundial, em 1996.

Por todas estas razões, o Porto apresenta-se como um misto de modernidade e conservadorismo, onde o passado convive com o presente, nas suas facetas mais diversas, incluindo as que dizem respeito às características particulares do seu modo de falar.

2.2.1 Alguns estudos sobre o falar do Porto

Como atrás referimos, o falar do Porto, fazendo parte do Dialecto Setentrional, tem sido sempre, pelos mais diversos dialectólogos, considerado um caso particular. Por essa razão, passamos a apresentar alguns estudos que a esse falar se referem.

Esta forma de falar, a variedade do Porto (VP), insere-se, de acordo com a denominação e divisão feita por Leite Vasconcelos no *Mapa Dialectológico* (1893-1897), no Dialecto Interamnense (dialecto Setentrional para Cintra). O dialectólogo destaca a peculiaridade dos traços fonéticos desta variedade. “A forte personalidade que atrás aludi, da região do Baixo Minho Douro Litoral (que tem como centro urbano mais importante – o Porto) aconselha que neste caso, a isolemos, embora dentro do grupo baixo-minhoto-duriense-beirão como zona bem individualizada, tomando como base para este isolamento,

por exemplo a ditongação, tão caracterizadora, das vogais tónicas fechadas [e] em [ej], [o] em [ow] (por vezes [w6]).” (Ibid. p.15).

Também Boléo (1961, p. 235), na sua divisão do Falar Minhoto, se refere a uma variedade específica, a do Porto: “ é nesta zona que se situa uma das cidades com mais “personalidade linguística – o Porto. O bem conhecido bairrismo dos seus habitantes manifesta-se também na linguagem, mantendo (mesmo as pessoas cultas) formas, pronúncias e sotaque tradicionais”.

Em síntese, e se tentarmos reunir o que ao longo dos estudos referidos tem sido dito acerca das características do falar do Porto, podemos destacar as seguintes:

- Em posição final, depois de ditongo ou consoante líquida [l] e [r], ocorre um fenómeno de paragoge, pela realização da vogal [ə].

- A vogal [o] sofre um fenómeno de ditongação, desde que ocorra em posição acentuada, independentemente do contexto fonético e da posição na palavra.

- Manutenção do ditongo [ow] em posição inicial, medial e final, independentemente da posição do acento.

- A vogal nasal [o~] sofre um fenómeno de ditongação, desde que ocorra em posição acentuada, em posição inicial e medial na palavra, independentemente do contexto fonético.

- As vogais [ɔ], [a] e [e] sofrem um fenómeno de ditongação, desde que ocorram em posição acentuada, antes das consoantes palatais [j], [ɹ] e [ʒ] e da semivogal [j].

- A consoante [v], independentemente do contexto fonético em que ocorre e da posição na palavra, sofre um processo de substituição pela oclusiva bilabial [b].

Embora estes fenómenos fonéticos se estendam a outros falares do distrito do Porto, indicados na figura 3, eles são, sem sombra de dúvida, ainda hoje, também características dos portuenses. Isto é, quando um habitante do Porto ou do seu distrito se desloca a outras zonas do país, sobretudo para sul, é, pelos seus traços linguísticos, imediatamente reconhecido como “falando à Porto” ou sendo do Porto, no mínimo, como sendo do Norte.



Figura 3- Mapa do Distrito do Porto

Uma vez enunciados os fenómenos fonéticos característicos desta variedade, há necessidade de testarmos a sua actualidade. De acordo com toda a informação recolhida nos estudos acima enunciados, tornou-se fundamental, antes de prosseguir com a criação da voz, fazer um estudo experimental sobre estes fenómenos.

2.3 O Falar do Porto: breve estudo sobre a situação actual

Para aferir a actualidade e pertinência dos fenómenos fonéticos referidos por outros autores, decidimos realizar um estudo de caso. Para isso, procedemos a uma gravação informal, assim como à audição informal dos portuenses. Depois de procedermos à segmentação e análise das gravações efectuadas, constatámos que os fenómenos e traços diferenciadores referidos como característicos do falar do Porto ainda se mantêm, tendo mesmo sido encontradas outras variantes fonéticas, o que justifica, não só o aprofundamento do seu estudo, mas também a criação de uma voz com as características deste falar.

Não sendo possível um estudo mais exaustivo, e não sendo esse o nosso principal objectivo, optámos por realizar o estudo de um caso exemplar. Após a gravação de uma conversa informal com uma informante natural do Porto, anotámos o que considerámos relevante, para a caracterização deste falar. É com base nessa recolha e análise que

passámos a apresentar alguns exemplos de variação fonética considerados neste breve estudo.

2.3.1 Traços diferenciadores

Com base nas gravações efectuadas, podemos então confirmar que os traços fonéticos diferenciadores do Falar do Porto são, de uma maneira geral, os fenómenos de ditongação - manutenção dos ditongos ou mesmo a ditongação de uma vogal -, bem como o fenómeno de substituição do fonema /v/ por /b/. Outros existem, como por exemplo a abertura e consequente recuo da vogal /a/, mas são menos evidentes e, por isso, menos referidos como característicos deste Falar.

2.3.1.1 O [ə]

Ao contrário do que é comumente aceite para a Língua Portuguesa, em que a vogal oral, fechada, central, alta [ə] não tem geralmente realização fonética em nenhum contexto do PE, facto esse comprovado pelos estudos de Gonçalves Viana (1973), Delgado Martins (2002) e Andrade (1994), constatámos que, na variante do Porto, ocorre a paragoge dessa mesma vogal, em posição final, antecedida de ditongo ou das consoantes líquidas [l] e [r].

Na figura 4, apresenta-se a informação relativa à forma de onda e o espectro resultante da realização da palavra *eu*, pela nossa informante.

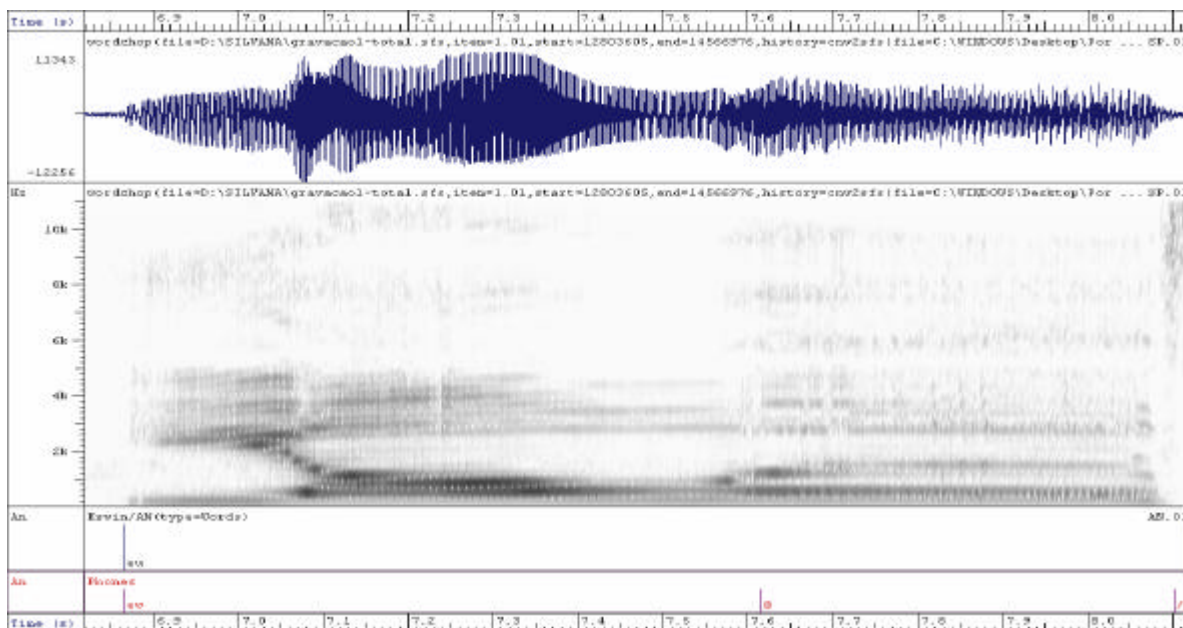


Figura 4 - Realização da palavra *eu* – eu[@]

Ao analisarmos esta figura, podemos constatar a realização do ditongo, neste caso [ew] seguido da vogal [@], bem visível no espectro, ocorrendo assim, a realização de [ew@] em vez de [ew].

Este tipo de fenómeno ocorre noutras palavras, como por exemplo: *mim* [mi~] realizado como [mi~J@]; *assim* [6si~] como [6si~J@]; *teu* [tew] como [tew@].

2.3.1.2 O [b]

No que diz respeito ao fenómeno de substituição do [v] pelo [b], no início de palavra, Paiva Boléo, no seu artigo *Dialectologia e História da Língua, Isoglossas portuguesas*, de 1951, caracteriza-o como “desaparecimento da oposição fonológica entre os fonemas /v/ e /b/, realizado ora como oclusiva, ora como fricativa (ou aspirante) b ou b’”. O autor considera este fenómeno característico do Falar Minhoto (correspondente ao Dialecto Setentrional), fazendo, também, alusão à sua existência na variedade do Porto. Podemos observar a ocorrência deste fenómeno de variação na figura 5, onde, a nossa informante, ao produzir a palavra *vida*, a realiza como [bida].

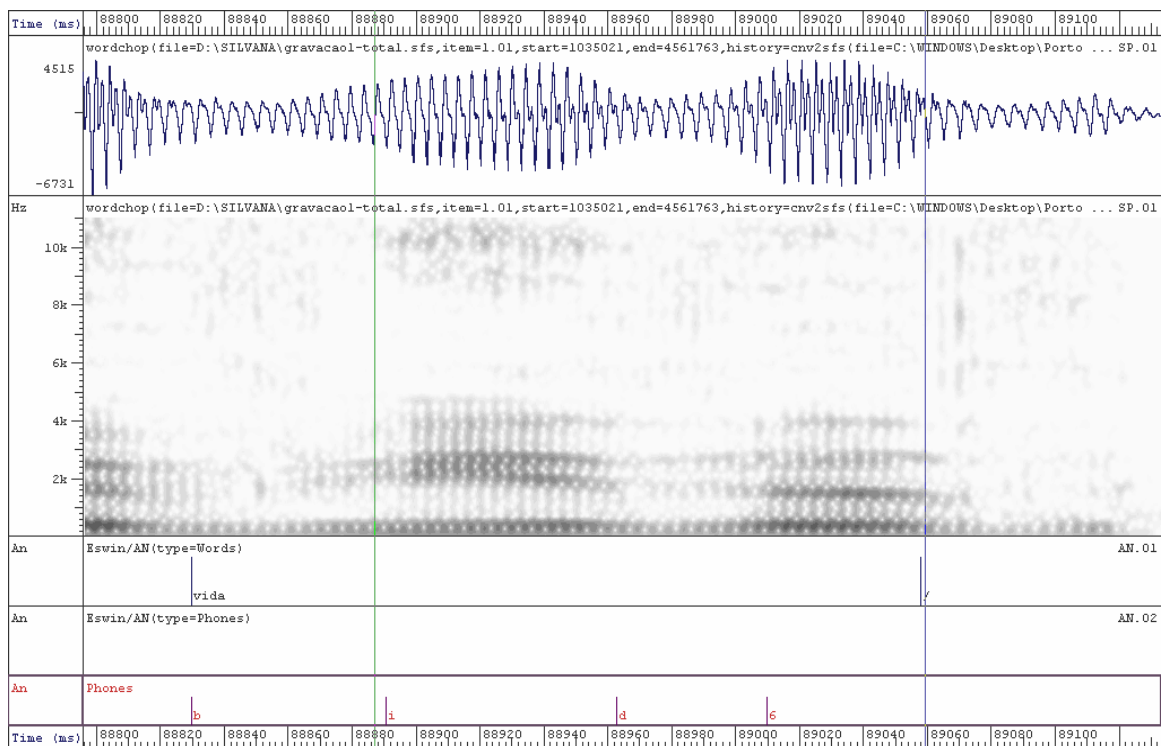


Figura 5 - Realização da palavra *vida* – [b]ida

O espectrograma, correspondente à realização da palavra *vida*, mostra claramente que a consoante inicial não se apresenta com características espectrais de uma fricativa. Com uma configuração muito próxima da consoante oclusiva sonora [d], constitui um exemplo de troca de [v] por [b], ou, segundo outros, da neutralização da oposição fonológica entre [v] e [b] no falar do Porto.

2.3.1.3 A ditongação

No que respeita ao número e ao tipo de ditongos existentes na Língua Portuguesa, as gramáticas e outros estudos que a eles se referem são, por vezes, lacunares e até descoincidentes. De uma maneira geral, consideram-se apenas, como ditongos da Língua Portuguesa, os ditongos decrescentes, orais e nasais. Quanto aos crescentes, existe consenso nos casos em que o [w] é precedido de [k] e [g]. Relativamente aos restantes, as opiniões divergem, sendo a sua realização considerada por uns e, se não negada por outros,

pelo menos omitida. No entanto, a presença destes últimos na Língua Portuguesa é sempre apresentada como excepcional, fazendo depender a sua realização de diversos factores, tais como o contexto fonético e a velocidade de elocução.

Na variedade em estudo, para além da manutenção da maior parte dos ditongos considerados na variedade normativa, e de acordo com os dados de que dispomos, ocorrem outros fenómenos de ditongação, que seguidamente descrevemos:

Para a vogal [o] confirma-se um fenómeno de ditongação, desde que ocorra em posição acentuada, independentemente do contexto fonético e da posição na palavra, realizando-se, frequentemente, como ditongo [w6]. São disso exemplo as seguintes realizações: [ˈnw6m@] *nome*, [ˈpw6rtu] *Porto*, [ˈtw6d6] *toda*, [ˈw6Z@] *hoje*, [6ˈbw6] *avô*. Como podemos observar pelos estudos realizados sobre o fenómeno de ditongação no PE, segundo a variedade normativa, nenhuma destas realizações é considerada, continuando a ser, no entanto, caracterizadora da variante em estudo.

A vogal nasal [o~] sofre também um fenómeno de ditongação, desde que ocorra em posição acentuada, em posição inicial ou medial, independentemente do contexto fonético. Ou seja, a falante em vez de realizar a vogal nasal [o~], produz a semivogal nasal [w~] e só depois a vogal nasal [6~], realizando [ˈw6~d@] *onde*, [ˈpw6~tu] *ponto*, em vez de [o~d@] e [po~tu], respectivamente.

As vogais [6], [a] e [e] sofrem um fenómeno de ditongação, desde que ocorram em posição acentuada, antes das consoantes palatais [j], [L] e [Z] e da semivogal palatal [j], como por exemplo: em [ˈmajL6] *malha*, [ˈpajL6] *palha*, [ˈsejJ6] *senha*, [ˈko~sejLu] *concelho*, [iˈgrejZ6] *igreja*, [6ˈrej6] *areia*. A estes tipos de realizações se refere já Moutinho (2001), aquando do seu estudo sobre fenómenos de ditongação na variedade do Porto.

Estes fenómenos ainda hoje se mantêm, como podemos constatar nas figuras 6 e 7, correspondentes a análises realizadas com base na recolha efectuada.

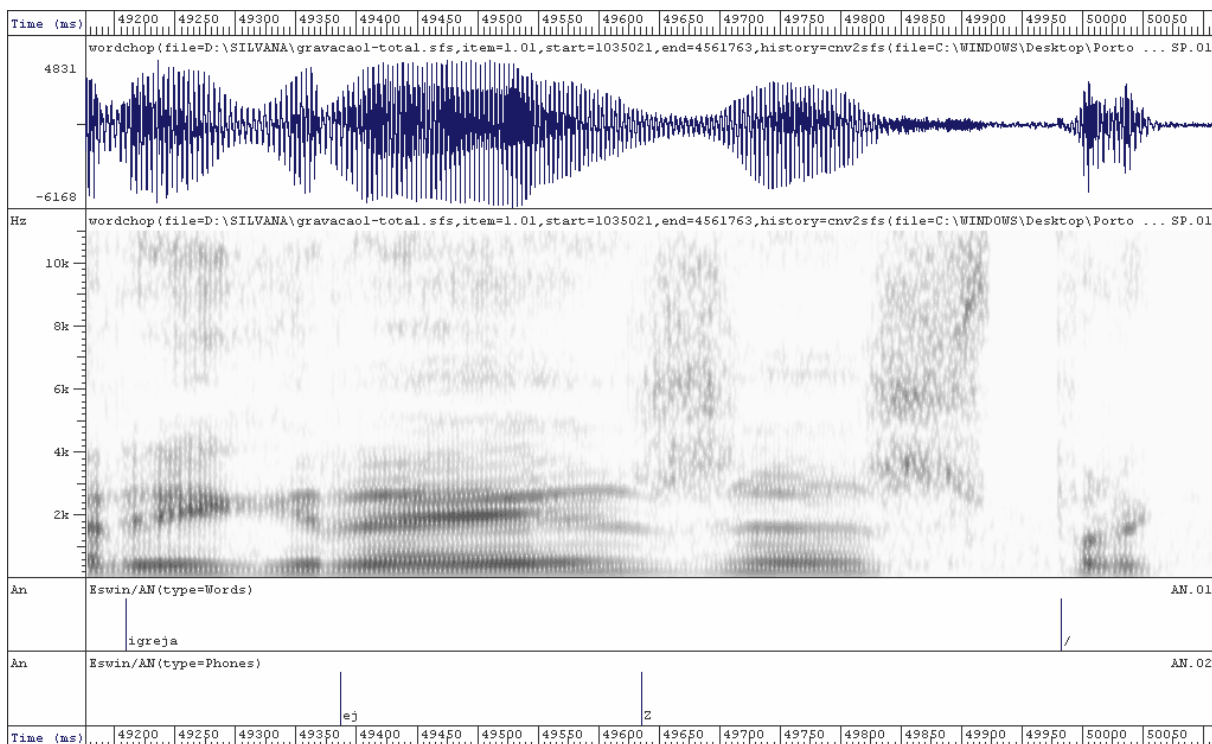


Figura 6 – Realização da palavra *igreja* – igr[ej]ja

A figura 6 atesta este fenómeno de ditongação, aquando da realização da palavra *igrejas*. Enquanto na variante normativa esta palavra se realiza como [i'gr6Z6S], podemos constatar, pela figura, que na variante do Porto se produz como [i'grejZ6S].

Verifica-se também a manutenção do ditongo [ow] em posição inicial, medial ou final, independentemente do acento, como por exemplo em ['owtru] *outro*, [ow'tonu] *outono*, ['powku] *pouco*, [Stow] *estou*. Este fenómeno é descrito já por Paiva Boléo no seu artigo *Dialectologia e História da Língua Isoglossas portuguesas*, de 1951, como traço característico do Falar Minhoto. Enquanto na variedade normativa, a palavra “outono” é produzida como [otonu], na variante do Porto teremos a realização [owtonu].

Para além disso, muitas vezes, a vogal [o] quando precede a vogal [6] sofre um fenómeno de ditongação – [ow6]. Este fenómeno resulta na criação do ditongo [ow], mesmo quando este não existe na grafia, como por exemplo na palavra *Lisboa* [liZbo6] que se realiza na variedade do Porto como [liZbow6].

O documento que reproduzimos na figura 7, correspondente à realização da palavra *pessoa*, atesta o que acabámos de referir.

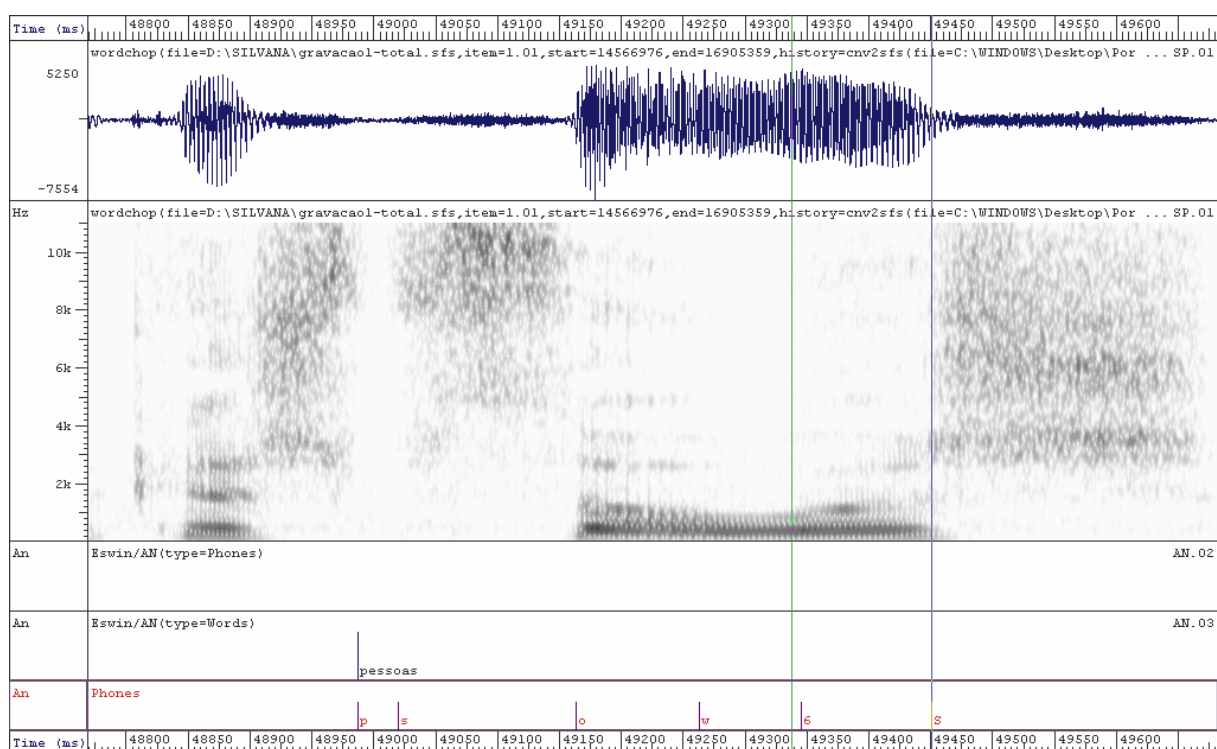


Figura 7 – Realização da palavra pessoa – pess[ow]a

Este breve estudo que acabámos de apresentar mostra a preservação dos fenómenos fonéticos já inventariadas como característicos da variante do Porto. Assim, parece-nos ser possível afirmar, ainda hoje, com Moutinho (2001, p.15), que “para os habitantes do Porto, estes traços de linguagem apresentam-se-lhes, quer como uma herança cultural a preservar, e da qual eles se sentem orgulhosos, quer como uma marca que os distingue do falar dos habitantes de outras regiões“, donde a pertinência e a actualidade da presente pesquisa.

3 Sobre Síntese e Sistemas de Síntese de Voz

Antes de entrarmos no capítulo referente à criação da voz do Porto, propriamente dita, (cap. IV), passamos a abordar aspectos relacionados com a síntese de voz e a sua evolução histórica, visto que o nosso objectivo, como foi já referido, passa pela construção de uma voz sintética.

A fala é o principal modo de comunicação entre as pessoas. A geração automática pelo computador de ondas de voz, conhecida como síntese de voz, tem merecido há várias décadas especial atenção por parte da comunidade académica, científica e empresarial.

Historicamente, a simulação das diferentes formas assumidas pelo canal bucal é o método mais antigo de síntese de voz. A primeira voz sintética foi criada para um autómato mecânico, reunindo um grupo de tubos e válvulas para simular o tracto vocal. Os modelos de computador que têm por base este processo não deram resultados convincentes até à data, visto serem modelos de extrema complexidade.

3.1.1 Síntese: o que é?

A síntese de voz pode ser definida, de uma forma muito geral, como a criação de voz através de dispositivos ou processos artificiais. Como, em geral, o interesse da síntese se centra na utilização da voz no âmbito da linguagem humana, é mais usual a utilização da designação síntese de fala.

Normalmente, a síntese está associada a sistemas que têm por entrada um texto passível de ser convertido em voz (em inglês, sistemas text-to-speech (TTS)). Estes sistemas são definidos por Dutoit (1997, p.13) como a produção de voz por máquinas através da fonetização automática do texto. O objectivo principal deste tipo de síntese consiste na capacidade do sintetizador ler um texto, de qualquer natureza, introduzido directamente no PC, ou submetido através da leitura óptica. A leitura do sintetizador pretende-se inteligível e natural.

O processamento nos sistemas TTS consiste em duas fases principais. A primeira é a análise do texto, onde o texto de entrada é convertido em fonemas ou outra representação linguística; a segunda é a geração de formas de onda da voz, onde a saída acústica é a produção destes fonemas e da informação prosódica. Estas duas fases são usualmente chamadas de síntese de alto e baixo nível, respectivamente.

3.2 Estrutura dos sistemas TTS

Apesar da sua diversidade, a maioria dos TTS apresenta uma estrutura comum. Podemos assim distinguir dois blocos principais: o de Processamento de Linguagem natural e o de Processamento de Sinal.

O primeiro bloco é passível de três subdivisões: Analisador de Texto – Text Analyser ou TA; Conversor de Texto para Fonemas – Text to Phoneme ou TTP; Gerador de Prosódia – Prosodic Generator ou PG.

O Analisador de Texto tem como principal função minimizar a variabilidade do texto de entrada. Assim, todos os caracteres especiais, como numerais (horas, datas, ordinais, cardinais e quantias em dinheiro), abreviaturas, acrónimos e símbolos vários têm de ser expandidos em palavras, em símbolos válidos e pertencentes ao alfabeto do idioma em estudo.

O Conversor de Texto para Fonemas converte o texto de entrada numa sucessão de fones, a partir dos quais é gerado o respectivo sinal. Contudo, esta tarefa é de extrema dificuldade, uma vez que não existe uma correspondência unívoca entre os grafemas e a sua realização oral.

O Gerador de Prosódia procura controlar algumas características da prosódia, temporais e tonais, para que a voz seja produzida da forma mais natural possível.

3.3 Um pouco de história

O primeiro trabalho em síntese de voz foi feito por volta de 1779 por Christian Gottlieb Krazenstein da Academia Imperial de St. Petersburg. Krazenstein inventou um instrumento que usava uma palheta vibrante e um constante fluxo de ar, como o mecanismo de um órgão (Von Mosser, 2000).

Mais tarde, em 1791, Wolfgang von Kempelen, de Viena, criou uma *máquina falante*, composta por um *fole* (pulmões) e uma *palheta* (cordas vocais). A forma da câmara de ressonância poderia ser alterada manualmente para gerar diferentes sons de voz, da mesma maneira que a posição da língua, dos lábios e maxilar alteram a forma do tracto vocal. Noutras palavras, estas máquinas foram baseadas na percepção de algumas características chave do aparelho de produção da voz (Sproat, 1998). Esta máquina conseguiu produzir não só sons de vogais, mas também palavras completas. Técnicos especiais foram treinados durante meses para usar a máquina e gerar fala inteligível.

Outros trabalhos no âmbito de *máquinas falantes* não ocorreram até 1939, até ao momento em que o VODER ou *Voice Operation Demonstrator* foi desenvolvido por Dudley, nos laboratórios *Bell*, para a Feira Mundial, na cidade de Nova York. Era constituído por dois geradores de sons independentes (ou excitação): um para os sons periódicos (cordas vocais durante sons vozeados) e outro para o ruído (turbulência causada pelas constricções no tracto vocal). Um filtro operado manualmente imitava os efeitos do tracto vocal. Pessoas foram treinadas para manusear o *Voder* para propósitos de demonstração. Este treino foi bastante longo. A máquina tinha um pedal de controlo da frequência manipulado pelo pé e 10 teclas para executar o sistema de controlo das ressonâncias.

O primeiro sintetizador de Formantes, PAT (*Parametric Artificial Talker*) foi apresentado por Walter Lawrence, em 1953 (Klatt 1987). Este sintetizador consistia em três formantes eléctricos ressoadores ligados em paralelo.

No que diz respeito à síntese articulatória, o primeiro aparelho surge em 1958, com George Rosen, do Instituto de Tecnologia de Massachusets, M.I.T. (Klatt 1987).

Em 1968, foi desenvolvido, no Japão, por Noriko Umeda, o primeiro sistema TTS completo para o Inglês.

A investigação sobre sintetizadores por concatenação de unidades surge nos anos 50/60. O primeiro sintetizador por concatenação, com base em difones, foi implementado em 1968, por Dixon e Maxey (Carvalho, 2004).

As técnicas de síntese, bem como a sua tecnologia, sofreram uma grande mudança nas últimas décadas. O aparecimento dos computadores digitais tornou possível simular sintetizadores. Os circuitos VLSI contribuíram, também, para esta mudança, tornando possível produzir um sintetizador de voz completo, ao utilizarem um circuito específico.

Nesta secção, apenas se apresentaram alguns momentos que considerámos mais importantes neste domínio. Muito mais informação pode ser obtida em Linggard (1985), por exemplo.

Os progressos recentes em síntese de voz têm sido motivados por três factores: o rápido aumento da capacidade dos computadores para realizar tarefas velozmente e com baixo custo; um grande aumento no número de texto e banco de vozes disponíveis; o aperfeiçoamento da tecnologia de reconhecimento de voz e de síntese (Dutoit, 1997).

Estes progressos têm originado sintetizadores com alta inteligibilidade, mas a qualidade do som e a naturalidade ainda continuam a ser um grande problema. Entretanto, a qualidade dos produtos tem alcançado um nível adequado para várias aplicações, tais como multimédia e telecomunicações. Com alguma informação audiovisual ou animação facial é possível aumentar consideravelmente a inteligibilidade da voz (Lemmetty, 2000).

3.4 Aplicações

Os sistemas de conversão de texto para voz abrangem numerosas e potenciais aplicações, das quais apresentaremos alguns exemplos.

A utilização destes sistemas nas telecomunicações possibilita uma maior oferta de serviços por parte destas empresas: voice-mail, sistemas de informações, sistemas de tradução automática, redes telefónicas inteligentes e sistemas multimédia são alguns exemplos.

O apoio à aprendizagem de uma nova Língua é também possível através destes sistemas. Estes permitem ao utilizador ver a palavra, o seu significado, bem como ouvir a sua realização. Desta forma, o utilizador terá acesso, não só à grafia da palavra, mas também aos sons que esta contém, facilitando-lhe a sua compreensão e assimilação. O ensino assistido por computador facilita aos alunos, que por alguma razão não podem frequentar a escola, o acesso à informação e às aulas.

O auxílio a indivíduos com deficiência é também uma aplicação muito importante destes sistemas. Este tipo de sistemas permite a um cidadão surdo-mudo comunicar com qualquer indivíduo, mesmo que este não conheça a Língua gestual. Os sistemas de conversão permitem, também, aos deficientes visuais, obter toda a informação necessária

para consultarem uma página da Internet, serem capazes de “ler/ouvir” um livro, um texto, ou mesmo apontamentos sobre uma determinada disciplina escolar.

A utilização destes sistemas para a construção de bonecos e livros “falantes” é uma área em grande desenvolvimento, tornando os brinquedos semelhantes ao ser humano, ou pelo menos com algumas características humanas.

Para além destas que acabámos de referir, não podemos deixar de realçar a utilização destes sistemas na investigação fundamental aplicada ao estudo das diversas Línguas. Os TTS podem ser uma ferramenta de grande utilidade para os linguistas, visto que lhes permitem realizar estudos experimentais sobre a Língua, possibilitando o aprofundamento do conhecimento, pela experimentação e validação de hipóteses. A validação de modelos prosódicos, das regras acústicas, dos métodos articulatórios e da existência de variedades linguísticas dentro de cada Língua, só é possível utilizando e testando estes sistemas de conversão.

3.5 Tipos de síntese

O bloco de Processamento de Sinal tem como principal componente um gerador de sinais de voz, também denominado de sintetizador. Este, por sua vez, permite dois tipos de síntese: síntese por regras (sintetizadores articulatórios e sintetizadores de formantes) e síntese por concatenação.

Os sintetizadores articulatórios procuram reproduzir os mecanismos fisiológicos de produção de voz e geração de sons pelo tracto vocal.

Nos sintetizadores de formantes, a produção é realizada, modelando as principais características acústicas do sinal de voz. O modelo acústico utilizado baseia-se na teoria da fonte/filtro.

Nos sintetizadores concatenativos utilizam-se diferentes exemplos (difones, fones, palavras,...) de voz natural, previamente gravados.

Os sistemas de síntese de formantes e os de concatenação são os mais utilizados na síntese de voz. Durante muito tempo, a preferência recaiu sobre os sintetizadores de formantes, no entanto, hoje em dia, são os concatenativos os mais utilizados.

O método articulatório continua a ser um método bastante complicado em termos da implementação e dos custos computacionais implicados. Contudo, será um método a desenvolver e aplicar no futuro.

3.6 Síntese por concatenação

Passamos em seguida a explicar o método de síntese por concatenação, em particular a concatenação de difones, visto ser o sistema que será utilizado nesta pesquisa para a criação da voz do Porto.

Os sintetizadores concatenativos produzem um sinal de voz, através da concatenação de segmentos de voz, previamente armazenados numa base de dados, obtidos por meio de gravações. Se bem que, para uma linguista como nós, fosse mais simples e mais óbvia a anotação de fones para a realização de síntese, preferimos adoptar a síntese por concatenação de difones. Esta metodologia permite-nos obter melhores resultados do que se tivéssemos optado pela utilização de fones. O grande problema é que, ao contrário da escrita, a produção de voz é um processo contínuo. A divisão desse contínuo nas zonas de transição entre fones torna muito difícil a sua posterior concatenação para criar uma voz sintética com um grau aceitável de naturalidade e inteligibilidade. Uma das soluções mais simples para esta questão é incluir a transição que ocorre entre os segmentos utilizados na concatenação, efectuando a junção em zonas mais estáveis, sem causar descontinuidade. Surgem assim os chamados difones, que se definem como os segmentos compreendidos entre a parte estável de um fone e a parte estável do fone que se lhe segue. Muitas vezes, por questões de implementação e de adequação a sons como as vogais, considera-se como parte estável o ponto médio do fone.

A utilização desta técnica implica um muito maior número de segmentos base do que os que seriam necessários, caso tivessem sido utilizados fones, visto que terão de ser contempladas todas as combinações possíveis de dois fones. Caso fossem possíveis, no PE, todas as combinações de fones, o número de segmentos/ difones seria igual ao quadrado do número de fones. Para o PE, passámos de cerca de quatro dezenas de fones para mais de milhar e meio de segmentos base.

Na fase de síntese, perante uma transcrição fonética, o sintetizador selecciona os segmentos adequados e concatena-os para gerar um sinal de voz. Por exemplo, na palavra

“pato” [´patu], distinguem-se os seguintes difones #p, pa, at, tu, u#, representando o sinal # o silêncio existente antes e depois da produção da sequência.

Na concatenação, os segmentos podem ser processados para suavizar as transições entre eles e para ajustar as estruturas prosódicas pré-definidas. (Black & Lanson, 2003).

A extração dos difones é baseada numa segmentação manual de cada um dos fones que compõe a sequência fonética, determinando-se automaticamente o ponto médio de cada fone, de maneira a encontrar o início e fim do difone (Black & Lanson, 2003, secção 11.4).

O resultado da concatenação depende, intrinsecamente, das regras de conversão grafema-fone que introduzimos, assim como da validação dos difones escolhidos, anotados e segmentados.

Em termos gerais, os sintetizadores concatenativos são os que oferecem melhores resultados, no que diz respeito ao processo de síntese.

3.7 Síntese do Português

De seguida, apresentamos os estudos em síntese de fala para o Português Europeu, que têm sido desenvolvidos por alguns grupos de investigadores.

Em 1991, surge o primeiro sistema de síntese de fala a partir de um texto desenvolvido de raiz para o Português Europeu - DIXI-, resultante da cooperação entre investigadores do grupo de Processamento de fala do INESC (Instituto de Engenharia e Sistemas de Computadores) e do CLUL (Centro de Linguística da Universidade de Lisboa). Este é um sintetizador por regra que utiliza o modelo de formantes Klatt.

Da cooperação entre o CEFAT (Centro de Estudos da Física Acústica e Telecomunicações, da Faculdade de Engenharia da Universidade do Porto) e a Universidade Técnica de Budapeste, surge, em 1996, o sistema Multivox para o Português, também ele, um sistema de síntese por regra com base em formantes.

Mais recentemente, foi apresentado na Faculdade de Engenharia da Universidade do Porto, um sistema de síntese por concatenação de difones (Barros 2002, citado por Carvalho 2004).

Nos sintetizadores articulatórios, destacam-se os trabalhos desenvolvidos na Universidade de Aveiro (Teixeira 2000), relativos à síntese dos sons nasais, e os realizados na Universidade do Minho (Silva 2001, citado por Carvalho 2004) para a estimação automática dos parâmetros do modelo articulatório.

As equipas de investigação do Laboratório de Sistemas de Língua Falada, conjuntamente com o Centro de Linguística da Universidade de Lisboa, estão a desenvolver uma nova versão do Sistema DIXI, mas agora, baseada em concatenação, no ambiente de desenvolvimento Festival, denominado DIXI+ . Este novo sistema tem como objectivo o desenvolvimento do sintetizador de fala, a partir de um texto para a comunicação alternativa e adaptativa (Carvalho, 2004). Uma das últimas evoluções deste sistema, realizada por Carvalho (2004) consistiu numa primeira abordagem à construção de um sistema de síntese de fala de domínio ilimitado, por concatenação de unidades de comprimento variável, para o Português Europeu.

3.8 O sistema Festival

Entre os vários sistemas usados para síntese concatenativa, encontra-se o sistema *Festival*, por nós adoptado neste trabalho.

Este sistema é dos mais populares, se não o mais popular, pelo seu carácter aberto, propiciador da criação de novas vozes em línguas em que já existe suporte e mesmo em novas línguas. Na sua estrutura interna, conta também com um suporte adequado a sistemas multilíngua e à realização de experiências em síntese por concatenação, pela disponibilização de scripts e manuais detalhados.

O *Festival* é um sistema multi-linguístico de síntese de voz, desenvolvido pelo *Centre for Speech Technology Research* (CSTR) da Universidade de Edimburgo, por Alan Black e Paul Taylor, em cooperação com CHATR, no Japão. O sistema está escrito em C++ e suporta LPC, os métodos PSOLA e as bases de dados MBROLA.

A sua arquitectura não difere muito da arquitectura geral de um qualquer TTS, uma vez que incorpora todos os seus módulos. Estes blocos oferecem um ambiente expansível, uma vez que é possível introduzir-lhes mudanças e/ou acrescentar novos blocos. Assim, os

utilizadores poderão adaptar o sistema às suas necessidades, desenvolvendo novas aplicações ou testando outras técnicas de síntese de voz. Trata-se de um sistema de utilização gratuita, com a vantagem de se poder trabalhar, tanto no sistema operativo *Windows*, como em *Linux*. O código da fonte pode ser alterado e personalizado, desde que se respeitem algumas regras de utilização pré-definidas.

Assim, e uma vez que o sistema dispõe de ferramentas que permitem a adição de um novo idioma, alguns dos módulos do *Festival* foram adaptados, tendo sido também criado um *corpus* de voz para o Português.

No âmbito do *Festival*, destaca-se ainda o projecto Festvox (<http://www.festvox.org>), que visa também tornar acessível a construção de novas vozes.

4 Construção de uma voz do Porto para o sistema *Festival*

Neste capítulo, descrevemos de forma sumária os vários passos da construção da voz do falar do Porto.

Uma vez que as vozes já criadas para o PE não são do domínio público, para a validação dos testes, foi necessário criar, através da gravação de outro informante, uma *voz standard*. Para a construção desta segunda voz, que procurámos que não apresentasse traços da variedade do Porto e se aproximasse o mais possível da variedade normativa, foram seguidos exactamente os mesmos passos que descrevemos para o FP. Desta forma, pudemos comparar e avaliar a importância da construção da voz do Porto, bem como o seu afastamento ou proximidade relativamente à *standard*.

4.1 Descrição Geral

Na construção das vozes seguiram-se as indicações e utilizaram-se as ferramentas disponibilizadas pelo projecto FESTVOX (Black & Lenzo, 2003), que tem como objectivo possibilitar e facilitar a criação de novas vozes para o sistema *Festival*.

De seguida, passamos a enunciar os vários processos necessários para a construção de uma base de difones, que serão descritos individualmente, numa fase posterior:

1. escolha da lista de fones
2. construção da lista de difones e respectiva lista de palavras onde eles ocorrem;
3. gravação do *corpus*;
4. anotação do *corpus*;
5. construção do ficheiro contendo informação relativa ao início e fim de cada um dos difones, assim como, o ficheiro de som em que se encontra armazenado (ficheiro wav.)
6. extracção das marcas de início e fim dos períodos de excitação glotal (pitchmarks);
7. extracção dos parâmetros necessários para a realização da síntese LPC (coeficientes e resíduos);

8. duração dos fones
9. teste e verificação da base de dados.

4.2 A escolha da lista de fones

O primeiro desafio na construção de uma voz consiste na definição dos fones necessários.

Na nossa lista, optámos por não incluir os alofones, como o [l~], nem os ditongos orais e nasais, de forma a reduzir o tamanho dessa mesma lista.

As maiores dificuldades prendem-se, neste caso, com a existência ou não de vogais nasais adicionais, o [E~] e [a~], referido por Sampson (1999). Se considerarmos os estudos dialectológicos realizados sobre as vogais nasais (Mateus, 1982; Barbosa, 1994; Teixeira et al., 2001) do Português Europeu, encontramos apenas cinco vogais nasais [6~], [e~], [i~], [o~] e [u~]. As vogais [a] e [E], na variedade normativa, não são realizadas como nasais. Optámos, no entanto, por incluir na nossa lista o fone [a~], visto este ser produzido por alguns falantes não escolarizados, durante as pré-gravações, em palavras como [ka~pu] *campo* e [ka~tu] *canto*. A vogal nasal [E~] não foi considerada como fone do PE, uma vez que não temos suporte textual ou experimental que confirme a sua realização.

Assim os fones considerados foram:

* Vogais(19):

- orais [a], [6], [@], [e], [E], [i], [o], [O] e [u]
- nasais:[a~], [6~], [e~], [i~], [o~] e [u~]
- semivogais: [j], [w], [j~] e [w~]

* Consoantes (19)

- Oclusivas:
 - orais sonoras: [b],[d],[g]
 - orais surdas: [p],[t],[k]
 - nasais: [m], [n], [J]
- fricativas:
 - orais: [v], [z] e [Z]
 - surdas: [f], [s], [S]

- laterais: [l] e [L]
- vibrantes: [r] e [R]
- o silêncio #

Assim, no total, a nossa lista de fones, em inglês “phoneset”, é constituída por 39 fones.

4.3 Construção da lista de palavras contendo os difones

Na nossa lista, tentámos incluir todas as sequências de fones passíveis de serem encontradas em palavras portuguesas, em sequências de palavras e em estrangeirismos realizados pelos falantes do Português Europeu.

Muitas vezes, opta-se por sequências de fones que não constituem palavra, para simplificar a tarefa de criação do *corpus*. No nosso caso, optámos por utilizar palavras, de forma a tornar a tarefa de elocução, pelo informante, mais natural e por ser, claramente, uma forma de incorporar conhecimentos linguísticos.

Tivemos a preocupação de tentar que todos os difones a anotar se encontrassem em posição medial, contudo nem sempre o objectivo foi conseguido.

A transcrição fonética da lista de palavras do nosso *corpus*, à semelhança do que já anteriormente referimos, foi realizada utilizando o alfabeto fonético SAMPA (Speech Assessment Methods Phonetic Alphabet), de forma a podermos usar o teclado comum do computador. As transcrições foram realizadas tendo como referência a variedade normativa. Apenas algumas sequências que nos pareceram ser mais específicas e pertinentes para o estudo foram anotadas de acordo com a variante do Porto.

Começámos por produzir todas as combinações possíveis de dois fones, incluindo o silêncio. Esta tarefa foi executada automaticamente pelo programa Diphone Studio, com base na lista de fones. De seguida, e por iterações sucessivas, foram procuradas palavras contendo as sequências. Depois de preenchidos os casos mais simples, com palavras comuns, seguiu-se a busca sistemática de sequências de palavras para associações de fones mais difíceis (ex: sequências vocálicas em hiato - VV).

No caso de consoante seguida de consoante, inicialmente optámos por admitir apenas os grupos consonânticos (oclusiva+líquida e fricativa+líquida) geralmente aceites (e.g. Lindley Cintra, 1997). Contudo, foi necessário analisar, também, outros encontros

consonantais menos frequentes, que surgem em palavras do tipo *abster*, *adventício*, *exportar*, *pneu*, *opção*, *adjectivo*, *adquirir*, *afta*, *istmo*, *etnografia*, *admoestar*, *magma*, *Magda*, *gnomo*, *amnístia*, *psicologia*.

De seguida, tratámos de eliminar todas as combinações CC sem realização no Português Europeu. A consoante líquida [R] só é produzida no início de palavra e em posição intervocálica. Eliminámos o [r], [J] e o [L], no início de palavra e/ou transição de palavras, assim como o [m], [n], [L], [J] e [R], no final de palavra e/ou sequências de palavras.

O mesmo procedimento foi seguido em relação às combinações vocálicas (VV), em sequências de palavras.

No intuito de tornar a lista o mais completa possível, decidimos incluir as diferentes formas de realização da mesma palavra. Assim, como a palavra “vida” foi produzida ora como [bid6] (pronúncia associada à variante do Porto) ora como [vid6] (segundo a variante normativa), ambas as formas foram consideradas. O nosso objectivo não é caracterizar apenas o nosso informante, mas produzir uma voz que seja reconhecida, por falantes oriundos de outras regiões, como uma voz característica do Porto, sem que, apesar disso, seja estereotipada.

Numa fase mais avançada, foram consideradas sequências apenas existentes em estrangeirismos. Foi feita uma pesquisa bastante minuciosa em dicionários de estrangeirismos (Machado, 1994 e Schmidt-Radefeldt, 1997) e nos recursos on-line disponibilizados pela Linguateca, (www.linguateca.pt), incluindo o *corpus* do *Público*. Com esta introdução, pretendia-se dotar a voz da capacidade de pronunciar estrangeirismos, pelo menos os resultantes dos recentes movimentos migratórios, muito visíveis na região do Porto.

O nosso *corpus* apresenta-se, assim, constituído por cerca de 1110 palavras, correspondentes ao mesmo número de difones, num total 1521 difones teoricamente possíveis. Ou seja, para algumas das combinações de fones geradas automaticamente pelo Diphone Studio não foram encontrados exemplos de ocorrência, pelo que estes difones foram considerados inexistentes no PE. Esta informação encontra-se condensada num ficheiro de texto, de forma a ser utilizável no processo de construção da voz para o Festival. Cada uma das linhas da lista corresponde a um difone e contém as seguintes

informações: fone inicial, fone final, nome para o ficheiro de som, transcrição fonética em SAMPA e ortografia.

i i raúl0.wav <viisu> vi isso

O nome do informante e o número do difone é usado para definir o ficheiro da onda sonora (wav.), o ficheiro lab. e outros ficheiros associados ao difone. Nas nossas listas fazemos a distinção do informante, com base no seu primeiro nome.

A parte relativa à ortografia é apresentada ao informante durante a gravação, constituindo o denominado *prompt*.

A lista anteriormente descrita é processada automaticamente, com vista à obtenção do seguinte formato, utilizável pelo Festival:

(raúl0.wav “pau v i i s u pau” (“i-i”))

isto é, nome do ficheiro – raúl0wav; transcrição com espaço entre os fones e pausas no início e fim da sequência (pau) - “pau v i i s u pau”; e o difone - (“i-i”).

4.4 Gravação do corpus

Após a construção da lista de palavras e selecção dos informantes, passámos à gravação do *corpus*.

O objectivo destas gravações é a obtenção de uma voz com qualidade, naturalidade e inteligibilidade. Assim, cada informante deve estar relaxado, à vontade, confortável, sem problemas de rouquidão ou qualquer outra limitação física que possa interferir na qualidade dos sons ou sequências a produzir.

O espaço da gravação deve ser, atempadamente, organizado, e ter o mínimo de barulho possível. O material deve ter sido testado previamente e estar pronto para utilizar. É importante reduzirmos ao máximo as probabilidades de interrupções e problemas de gravação.

Optámos por apresentar as palavras a realizar num programa de tratamento de sinal acústico – EMU. Depois de as visualizar no monitor, o informante teria de produzir a

palavra e/ou sequências de palavras. O sistema EMU (<http://emu.sourceforge.net/>), tem a vantagem de facilitar as tarefas de gravação e anotação, permitindo o acesso, ao mesmo tempo, ao espectograma e ao oscilograma.

Na escolha dos nossos informantes tivemos em conta as recomendações do manual do *Festival*.

Para a gravação foram seleccionados dois informantes: um, com realizações fonéticas características do falar do Porto (Raul); um outro (Luís), representante da variedade normativa.

Não sendo, de modo nenhum, nossa intenção criar uma voz estereotipada, como já tivemos ocasião de referir anteriormente, optámos por seleccionar para a voz do Porto, um informante do sexo masculino, com cerca de quarenta anos e formação superior, mas que mantém, apesar de tudo, traços característicos do FP, reconhecidos pelo próprio e por terceiros.

O informante, para a voz standard, é também do sexo masculino, com cerca de 30 anos e formação universitária.

A gravação foi feita em duas tardes, uma para cada informante, no Laboratório de Fonética da Universidade de Aveiro. Os informantes visualizaram no monitor do computador as palavras e ou sequências de palavras que, de seguida, reproduziram. No decorrer da gravação, foram realizadas apenas curtas pausas, para que o ritmo de produção e a prosódia não sofressem muitas alterações. O equipamento utilizado foi o Kay CSL4400 com gravação directa para o disco do PC.

4.5 Anotação do corpus

Nesta fase do trabalho, optámos por anotar os fones em causa e o contexto de ocorrência, de forma a obter o máximo de informação e o máximo de exemplos dos difones existentes. Incluímos também a pausa inicial ou final, sempre que o difone se encontrava em início ou fim de palavra.

A anotação foi feita inicialmente de forma automática, através do comando *phonealign*, do programa *Edimburg Speech Tools*. Embora a anotação tenha sido efectuada

de forma automática, posteriormente, foi sujeita a uma correcção manual, de forma a eliminar alguns erros da anotação automática.

A síntese dos *prompts*, necessária para o processo de anotação automática, foi feita apenas para a anotação, usando uma tabela que transforma um fone no outro, fazendo a correspondência entre o sistema, já existente, para o inglês e os fones que aí ocorrem e os que se realizam no português.

Os *prompts* são gerados com base na lista de difones e fazendo correr o comando

```
Festival festvox/us.chema.scm festvox/diphlist.scm
```

```
Festival>(diphone_gen_shema "us" etc/usdiph.list")
```

As ondas sonoras das *prompts*, resultantes da síntese, são guardadas no ficheiro *prompt*.wav* e as etiquetas (labels), conseguidas através da anotação, ficam registadas nos ficheiros com o formato *prompt*.lab*.

Apresentamos, na figura 8, um exemplo de anotação para a sequência “*pé positivo*”.

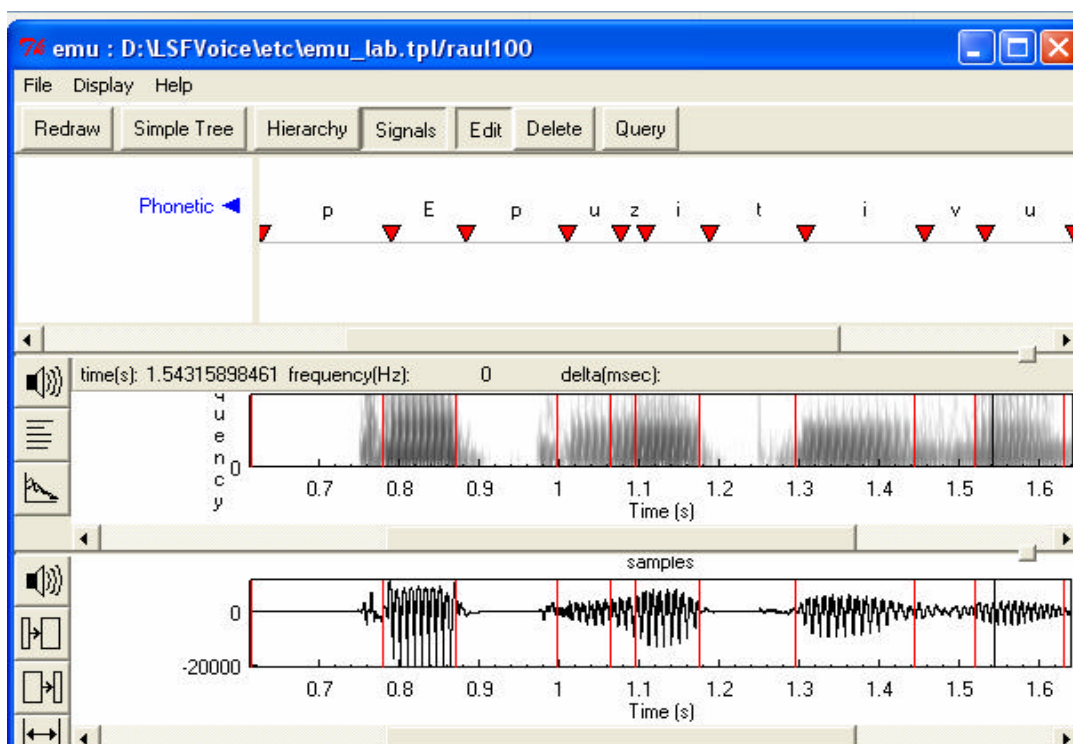


Figura 8 - Anotação da sequência *pé positivo*.

4.6 Construção do índice de difones

O passo seguinte é a criação de um índice onde podemos encontrar a informação sobre cada difone, nomeadamente a sua designação (fone inicial+hífen+fone final), nome do ficheiro, indicações sobre o início, a fronteira e o final do difone.

Esta informação foi extraída automaticamente através do comando:

make_diph_index

obtendo-se a seguinte informação:

```
EST_File index
DataType ascii
NumEntries 1110
EST_Header_End
pau-R Raúl1596 0.0000 0.00000 0.00000
pau-J Raúl1596 0.3175 0.635 0.73
pau-n Raúl1595 0.232781 0.465561 0.559265
pau-m Raúl1594 0.205455 0.41091 0.490513
pau-L Raúl1786 0.00000 0.0000 0.00000
....
```

Os problemas ocorridos em alguns difones são identificados pela presença de zeros nas três colunas à direita. Estes foram imediatamente corrigidos, antes de continuar o processo. Muitas vezes, a correcção passou pela alteração da anotação e/ou transcrição.

4.7 Extracção da informação dos períodos glotais e parâmetros LPC

Aquilo que é de facto utilizado na síntese não é o sinal de voz directamente, mas sim um processo de síntese designado por síntese LPC. Mais detalhes sobre esta questão estão fora do âmbito desta dissertação, sendo apenas referidos, de seguida, os comandos utilizados para a conclusão da criação de voz.

A informação sobre os períodos de abertura das cordas vocais (*pitchmarks*, em inglês) foram obtidos, utilizando o comando:

```
bin/make_pm_wave wav/*.wav
```

Os resultados foram corrigidos através do comando

```
bin/make_pm.fix pm/*.pm
```

Por forma a minimizar as diferentes amplitudes dos vários fones, foram de seguida calculados factores de correcção com o comando

```
bin/findpowerfactors lab/*.lab
```

Finalmente, utilizámos o comando *make_lpc* para obter os coeficientes LPC e os resíduos necessários para a síntese.

4.8 Duração dos fones

Com base nas anotações, foram calculados os valores médios de duração para todos os fones, tendo esses valores sido integrados no processamento da duração da voz por nós criada no sistema *Festival*. Este processamento apenas foi realizado para a voz do Porto, sendo os seus valores usados também para o falante da voz *standard*. Este processo permitiu atenuar a artificialidade decorrente da ausência de parâmetros prosódicos.

4.9 Teste e verificação da base de dados

Concluído o processo de criação da voz, efectuámos um primeiro teste às vozes obtidas, que consistiu na síntese de todas as palavras e sequências da nossa lista com base na sua transcrição, usando o comando *Festival SayPhones*. Nesta fase, foram detectadas e corrigidas algumas omissões na lista de difones, assim como alguns erros de anotação e mesmo de transcrição, naturais, dado o número elevado de difones.

Durante a fase de síntese, foram detectadas algumas lacunas relacionadas com o inventário de difones. Numa primeira etapa, não fomos capazes de encontrar exemplos de palavras ou sequências onde ocorressem determinadas combinações de fones, pelo que estes não foram incluídos na lista de possíveis difones do PE (e.g. difone [no~] existente na palavra *nongentésimo*). No entanto, este problema foi facilmente corrigido, uma vez que,

ao anotarmos não só os difones em causa, mas também a sequência restante, foi possível obter mais informação sobre determinadas sequências de fones que, inicialmente, não tinham sido contemplados.

Foram detectados outros problemas relacionados com três tipos de factores:

- ? inconsistência no tratamento dado aos ditongos, uma vez que não são contemplados no âmbito do nosso estudo;
- ? ocorrência de outras consoantes em coda, para além das tradicionalmente inventariadas para o PE - [r], [l] e [S] –, nomeadamente em estrangeirismos (e.g. *ketchup*);
- ? falhas nas sequências de consoante seguida de semivogal.

Assim, só nesta fase, foram incluídos os seguintes difones: [tw], [kw], [Ja~], [J6~], [v#], [Z#], [kj], [tj] e [SO]. Todos eles foram retirados de palavras e ou sequências integrantes da nossa lista.

No total, foram encontrados e corrigidos 28 erros em ambos os informantes.

Para teste, foram também sintetizadas outras pequenas sequências, com pronúncia standard e pronúncia da variante do Porto. A título de exemplo, para a sequência “a porta azul”, obtivemos [6pOrtazul] – pronúncia standard; e [6pwOrtazul@] – realização da variante do Porto.

Nesta fase, quer o material integrante da lista, quer as pequenas frases foram objecto de uma avaliação informal, tendo-se corrigido alguns erros de anotação, ou mesmo de processamento.

5 Resultados e discussão

O objectivo deste trabalho é a criação das vozes, que constituem em si mesmas a nossa ferramenta de trabalho. Por esta razão, revela-se de grande importância avaliar a sua qualidade.

Neste capítulo, procedeu-se a uma avaliação da qualidade/ inteligibilidade das duas vozes.

Numa primeira fase, foi efectuado um teste de percepção, com o objectivo de verificar se estas são ou não inteligíveis para os ouvintes .

Num segundo momento, procedemos a uma avaliação das vozes criadas, testando se estas eram ou não reconhecidas como do Porto. Tanto para a voz standard como para a do Porto, foram tidas em conta regras de conversão grafema-fone.

A avaliação incide exclusivamente sobre as vozes. Não nos podemos esquecer que estas não contêm informação prosódica, sendo também todos os sons produzidos com a mesma duração, o que as torna pouco naturais e de difícil percepção num primeiro momento.

5.1 A avaliação das vozes

Descrevemos, de seguida, o teste referente à inteligibilidade da voz.

5.1.1 Procedimento

Para obter uma medida de qualidade com significado, a avaliação dessa mesma qualidade deverá basear-se na resposta dos ouvintes (Quackenbush et al., 1988, citado Teixeira, 2000).

Numa primeira fase da avaliação, optámos por um teste de identificação (Sproat, 1988; Teixeira e Vaz 2000), usando o material existente na nossa lista, ou seja, as palavras e sequências de palavras que gravámos e anotámos.

No sentido de obtermos informação acerca da percepção e da inteligibilidade de cada um dos estímulos apresentados a cada ouvinte, foi-lhes solicitado que ouvissem as produções e, de seguida, as reproduzissem, mesmo que lhes parecessem estranhas ou impossíveis. Durante a realização do teste, foi-lhes igualmente pedido que classificassem quantitativamente, quanto à sua qualidade, as realizações ouvidas. Os estímulos deveriam ser avaliados numa escala de um a três, sendo o “um” correspondente a uma realização de *má qualidade*, o “dois” a uma produção de *qualidade satisfatória* e o “três” a um estímulo de *boa qualidade*. Deste modo, ao mesmo tempo que executávamos um teste de percepção que nos permitia averiguar a inteligibilidade da voz, também pudemos realizar, em simultâneo, um teste de qualidade.

O teste foi realizado, numa primeira fase, com a voz standard, sendo de seguida efectuado com a voz do Porto.

Foram seleccionados automaticamente e de forma aleatória cinquenta estímulos diferentes a apresentar a cada um dos cinco ouvintes, o que perfaz um total de duzentas e cinquenta avaliações (50 estímulosx5 ouvintes). O teste foi realizado individualmente, numa sala com ruído baixo a moderado, através de auscultadores.

Todos os informantes desconheciam completamente a lista, tendo apenas informação acerca do número de estímulos. Algumas das sequências de palavras apresentadas revelam baixa probabilidade de realização, uma vez que, na Língua Portuguesa, constituem sequências sem sentido. Estas foram incluídas pelo facto de só desta forma podermos contemplar todos os difones do PE.

Na realização dos dois testes – voz standard e voz do Porto - utilizámos os mesmos estímulos e os mesmos ouvintes. Para evitar que um ouvinte memorizasse os estímulos, quando solicitado a ouvir a segunda voz – a do Porto –, nunca lhe foi dada a mesma sequência de estímulos com que tinha sido confrontado anteriormente. Evitou-se, assim, o falseamento dos resultados pelo conhecimento prévio dos cinquenta estímulos.

5.1.2 Estímulos

Foi desenvolvido um pequeno programa, usando a linguagem em ambiente Matlab, para permitir a selecção automática e arbitrária de cinquenta estímulos resultantes da síntese da nossa lista de palavras e sequências.

Tivemos a preocupação de que cada ouvinte ouvisse cinquenta estímulos diferentes, de forma a obtermos o máximo de informação possível sobre a nossa lista, sem tornar o teste demasiado moroso e cansativo. Os estímulos, ou seja a tabela de palavras e sequências de palavras seleccionadas pelo computador, encontram-se em anexo (anexo 2,3).

5.1.3 Ouvintes

Os ouvintes foram escolhidos tendo em conta os seguintes factores: faixa etária e nível de escolaridade. Um grupo de ouvintes situa-se numa faixa etária compreendida entre os 20 e os 30 anos de idade, com formação superior; o outro grupo tem idades compreendidas entre os 40 e 50 anos, possuindo apenas o 6º ano de escolaridade. Participaram nos testes 5 ouvintes, sendo um do sexo masculino e os restantes quatro do sexo feminino, todos falantes nativos do Português Europeu, naturais e residentes, na sua maioria, na zona litoral norte do país. Assumidamente, todos eles têm contacto e conhecimento das duas variantes em análise.

Os ouvintes da voz standard são tratados como O1, O2, O3, O4 e O5. Os da voz do Porto são designados por OP1, OP2, OP3, OP4, OP5.

Tabela 1- Os ouvintes do teste de percepção

Nome	Sexo	Idade	Escolaridade	Residência
O1 OP5	Feminino	28	Licenciada	Macedo de Cavaleiros
O2 OP1	Feminino	25	Licenciada	Sta. Maria da Feira
O3 OP4	Feminino	23	licenciada	Sta. Maria da Feira
O4 OP2	Masculino	48	2º Ciclo do E.B	Sta. Maria da Feira
O5 OP3	Feminino	45	3ºCiclo do E.B	Sta. Maria da Feira

5.1.4 Resultados para a voz com pronúncia standard

Os resultados dos testes foram classificados de acordo com as categorias que de seguida passamos a enunciar: *acertou*, *acrescento de som inicial*, *desnasalização*, *ditongação*, *elisão do som inicial*, *semivocalização*, *só entendeu a segunda parte*, *só entendeu o fim*, *só entendeu o som inicial*, *sonoridade próxima*, *troca de fricativas*, *troca de oclusivas*, *troca de vogais*, *troca fricativa vibrante e troca oclusiva vibrante*, *não entendeu*.

De seguida, passamos a descrever o que entendemos por cada uma das classificações.

Acertou: esta classificação foi utilizada sempre que o ouvinte acertou completamente na palavra e/ou na sequência apresentada.

Acrescento de som inicial: este fenómeno, designado por prótese, surge em palavras como *arfar* [6rfar], reproduzida pelo ouvinte como [f6rfar], com acrescento, neste caso, do [f] inicial.

Desnasalização: em alguns casos a nasalidade não é percebida durante a audição. Exemplos desse fenómeno são as sequências *som antigo* [so~6~tigu], *som inteiro* [so~i~tejru], e *som entre* [so~e~tr@] que são entendidos como [sO6~tigu], [sOi~tejru] e [sOe~tr@], respectivamente. A palavra *manhã* [m6J6~] é também exemplo deste fenómeno, sendo entendida como [maJ6].

Ditongação: a vogal [O] sofre um fenómeno de ditongação, realizando-se como [aw] na sequência *pó inteiro* [pOi~tejru] que é reproduzida pelo ouvinte como *pau inteiro* [pawi~tejru]. Este fenómeno pode ficar a dever-se ao facto da segunda sequência *pau inteiro* ser mais provável na Língua Portuguesa e fazer mais sentido do que a primeira *pó inteiro*.

Elisão do som inicial: algumas respostas dadas pelos ouvintes apresentam um fenómeno denominado de aférese. São disso exemplo *cantado* [k6~tadu], que é ouvido como [6~tadu], e *cândido* [k6~didu], entendido como [6~didu], em que ocorre a queda do [k]; em *pó* [pO] e *pé positivo* [pEpuzitivu] o [p] inicial desaparece, sendo os estímulos entendidos como [O] e [Epuzitivu], respectivamente; e o som [t] que cai na palavra *tanto* [t6~tu], referida como [6~tu].

Semivocalização: este caso ocorre apenas uma vez - a consoante [l] é entendida como a semivogal [w], daí que *malzinho* [malziJu] tenha sido percebido e realizado como *mauzinho* [mawziJu].

Nestes três casos, que de seguida apresentamos, o ouvinte apenas percebeu uma parte da palavra ou sequência de palavras: umas vezes **só entendeu a segunda parte**, como nas sequências *pó ondulado* [pOo~duladu] e *pau nhónhó* [pawJOJO], reproduzidas como *ondulado* [o~duladu] e *dá-lhe nhónhó* [daL@JOJO]. Outras vezes o ouvinte **só entendeu o fim** como acontece nas sequências *pau pequeno* [pawp@kenu] e *de Rita* [d@Rit6] identificando-as como *albequeno* [6lb@kenu] e *visita* [vizita]. Outro dos fenómenos é exactamente o oposto, já que o ouvinte **só entendeu o som inicial**. Assim, na sequência *som antigo* [so~6~tigu], apenas ouviu [so~].

Queda do som [l] em posição medial: a palavra *polvo* [polvu] sofre uma síncope da consoante líquida [l], uma vez que esta é entendida como *povo* [povu]. O informante não realizou o [l] medial. Este fenómeno verifica-se apenas no teste com a voz do Porto.

A classificação **sonoridade próxima** é utilizada para casos em que o ouvinte não entende a palavra e/ou sequência, mas reproduz algo de semelhante a nível sonoro, tal como acontece em *pó inchado* [pOi~Sadu], em vez de *põe chave* [po~j~Sav@], *amor* [6mor] por *tambor* [t6~bor], *pó impado* [pOi~padu] por *pó untado* [pOu~tadu], *vê areal* [ve6rial] por *pé ideal* [pEidial], *esta a sentir* [St66se~tir] por *assentir* [6se~tir], *cedo* [sedu] por *céu* [sEw], *manhã sem enchada* [m6J6~s6~j~e~Sad6] por *manhã fechada* [m6J6~f@ Sad6], entre outros.

Troca de fricativas: [s] realiza-se como [S] na palavra *ranço* [r6~su] que se transforma em *rancho* [r6~Su]; [Z] realiza-se como [s] na sequência *com jeito* [ko~Zejtú], substituída por *conceito* [ko~sejtú]; [z] realiza-se como [v] no caso de *zinco* [zi~ku] que passa a *vinco* [vi~ku]; o [f] realiza-se como [S] em *kafka* [kafk6] passando a *casca* [kaSk6].

Troca de oclusivas: [p] pelo [b] na palavra *bato* [batu] produzida como *pato* [patu]; [t] pelo [d] em *doente* [due~t@] que passa a *duende* [due~d@]; [d] pelo [p] na sequência *dai novo* [dajnovu] em vez de *pai novo* [pajnovu] e na palavra *golde* [gOld@] em vez de *golpe* [gOlP@]; o [b] pelo [d], onde o ouvinte refere *be* [b@] quando o estímulo é *de* [d@].

Troca da consoante fricativa por vibrante: o ouvinte identifica a vibrante [r] na realização da fricativa [v], pois identifica *avô único* [6vouniku] como *aro único* [aruuniku].

Troca da consoante oclusiva por vibrante: *agitar* [6Zitar] é substituída por *a girar* [6Zirar] e concretiza a mudança da consoante oclusiva [t] para a consoante vibrante [r].

Troca da consoante oclusiva por fricativa: a sequência *põe feliz* [po~j~f@liS] é reproduzida como *põe veliz* [po~j~v@liS]. Isto é, a consoante oclusiva [p] é identificada, neste exemplo, como a consoante fricativa [v] (fenómeno verificado apenas no teste com a voz do Porto).

Troca de vogais: este fenómeno verifica-se em exemplos como: *como antiga* [komu6~tig6] que foi ouvida *como antigo* [komu6~tigu]; *manhã nova* [m6J6~nOv6] substituída por *manhã novo* [m6J6~novu]; *de antiga* [d@6~tig6] entendida como *de antigo* [d@6~tigu], onde a vogal [6] passa a [u]. Também as sequências como *vê antigo* [ve6~tigu] e *manhã esse* [m6J6~es@] se vêem substituídas pelas sequências *vi antigo* [vi6~tigu] e *manhã isso* [m6J6~isu], respectivamente, onde a vogal [e] é percebida e realizada como [i] .

Não entendeu: nesta categoria, incluímos todas as palavras e sequências de palavras que o ouvinte não conseguiu de todo entender, não sendo, por isso, capaz de as reproduzir . São disso exemplo os seguintes itens: *são-lhe* [s6~w~L@], *polvo* [polvu], *alfândega* [alf6~d@g6], *zero* [zEru], *atum jeitoso* [6tu~Z6jtozu], *ramo* [r6mu], *pedinte* [p@di~t@], *atum entre* [6tu~e~tr@], *pai esse* [pajes@], *são lindos* [s6~w~li~duS], *exame* [e~S6m@], *pau só* [pawsO], *prenda* [pre~d6] e *amnistia* [6mniSti6].

De acordo com a avaliação dos testes, obtivemos os resultados globais representados graficamente na figura 9.: Ao analisarmos a figura, podemos verificar que a percentagem de acerto é elevada (a azul). Também com percentagens significativas, destacamos a ocorrência de casos de sonoridade próxima (laranja) e também estímulos não entendidos (vermelho).

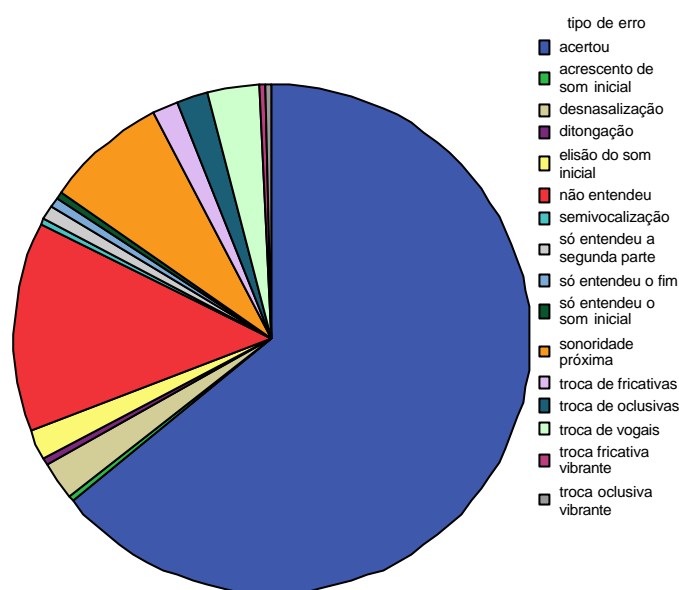


Figura 9- Resultados globais do teste de percepção para a voz standard

Para uma análise mais detalhada, observem-se os valores, agora apresentados em tabela.

Tabela 2- Resultados globais e percentuais do teste de percepção para a voz standard

		Tipo de estímulo					
		Palavra		Sequência		sequência improvável	
		Casos	% palavras	Casos	% sequências	Casos	% seq improvável
tipo de erro	acertou	122	71,8%	19	59,4%	19	39,6%
	acrescento de som inicial	1	,6%	0	,0%	0	,0%
	desnasalização	2	1,2%	0	,0%	4	8,3%
	ditongação	0	,0%	0	,0%	1	2,1%
	elisão do som inicial	4	2,4%	0	,0%	1	2,1%
	não entendeu	20	11,8%	6	18,8%	7	14,6%
	semivocalização	1	,6%	0	,0%	0	,0%
	só entendeu a segunda parte	0	,0%	0	,0%	2	4,2%
	só entendeu o fim	0	,0%	1	3,1%	1	2,1%
	só entendeu o som inicial	0	,0%	0	,0%	1	2,1%
	sonoridade próxima	9	5,3%	1	3,1%	9	18,8%
	troca de fricativas	4	2,4%	0	,0%	0	,0%
	troca de oclusivas	4	2,4%	1	3,1%	0	,0%
	troca de vogais	2	1,2%	3	9,4%	3	6,3%
	troca fricativa/ vibrante	0	,0%	1	3,1%	0	,0%
	troca oclusiva/ vibrante	1	,6%	0	,0%	0	,0%

Os resultados obtidos permitem-nos ter uma percepção quantitativa dos acertos e dos erros cometidos pelos ouvintes. A tabela contém os três tipos de estímulos apresentados aos ouvintes: palavra, sequência e sequência improvável. O estímulo que denominamos por “palavra” inclui palavras que ocorrem na Língua Portuguesa, bem como alguns estrangeirismos; por “sequência” entendemos uma sequência fonética onde se agrupam duas ou mais palavras da Língua Portuguesa; designamos por “sequência improvável” também uma sequência fonética, mas, neste caso, apenas constituída por duas palavras. No entanto, só quando estas são usadas isoladamente são reconhecidas como portuguesas, mas, quando agrupadas, originam uma sequência estranha e, normalmente, não utilizada em PE. Dentro de cada tipo dividimos a informação em número de casos ocorridos e a sua percentagem.

A percentagem de acerto na palavra foi de 71%, sendo os restantes 29% representantes dos tipos de erros ocorridos. As sequências obtiveram um grau de acerto de cerca de 59,4% e 40,6% de erros. Nas sequências improváveis, devido ao seu grau de dificuldade, o grau de acerto foi menor: cerca de 39,6% de acerto e 60,4% de erros.

A distribuição dos erros por cada tipo de estímulo (palavra, sequência e sequência improvável) é apresentada nos gráficos seguintes:

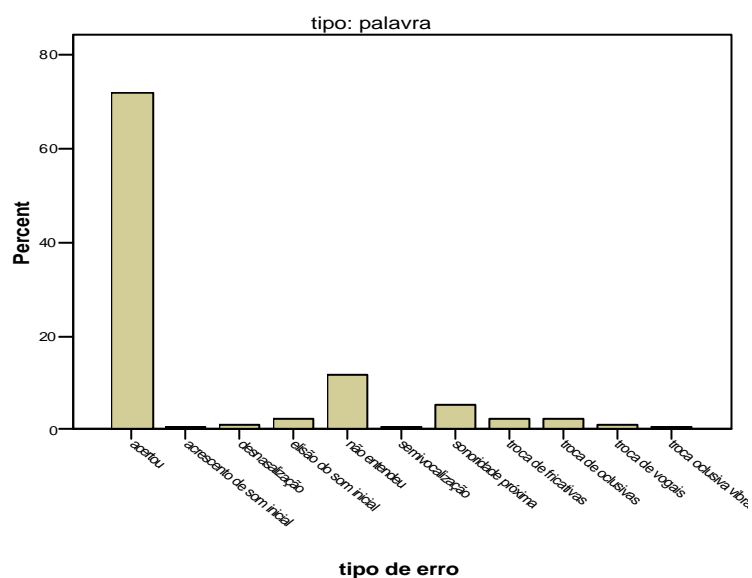


Figura 10- Ocorrência de erros no estímulo - palavra

Na figura 10, encontramos a percentagem de erros, no que diz respeito ao estímulo palavra. Como se pode verificar, a taxa de acerto atinge mais de 70%. Os erros com maior incidência referem-se a palavras não entendidas, seguidas das palavras que apesar de não terem sido percebidas completamente, são reproduzidas com sonoridade próxima.

Os restantes tipos de erros apresentam uma percentagem pouco significativa.

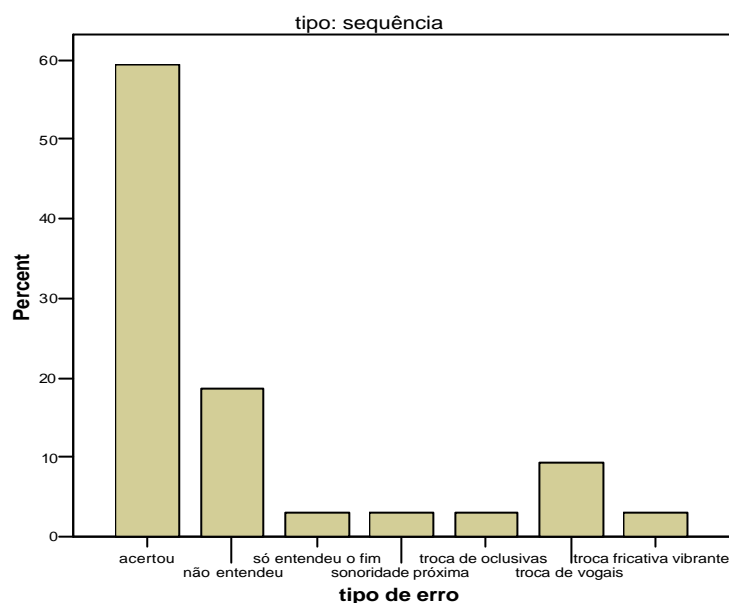


Figura 11- Ocorrência de erros no estímulo - sequência

Em relação às sequências, a figura 11 apresenta resultados de alguma forma divergentes dos anteriores. A percentagem de acerto é significativamente inferior, cerca de 60%, sendo a taxa de sequências não entendidas de 20%. A troca de vogais é também elevada.

Outro elemento a considerar é o número de categorias onde o erro ocorre : enquanto que, para as palavras, se realizam onze tipos de erros diferentes, nas sequências, o erro preenche apenas sete categorias.

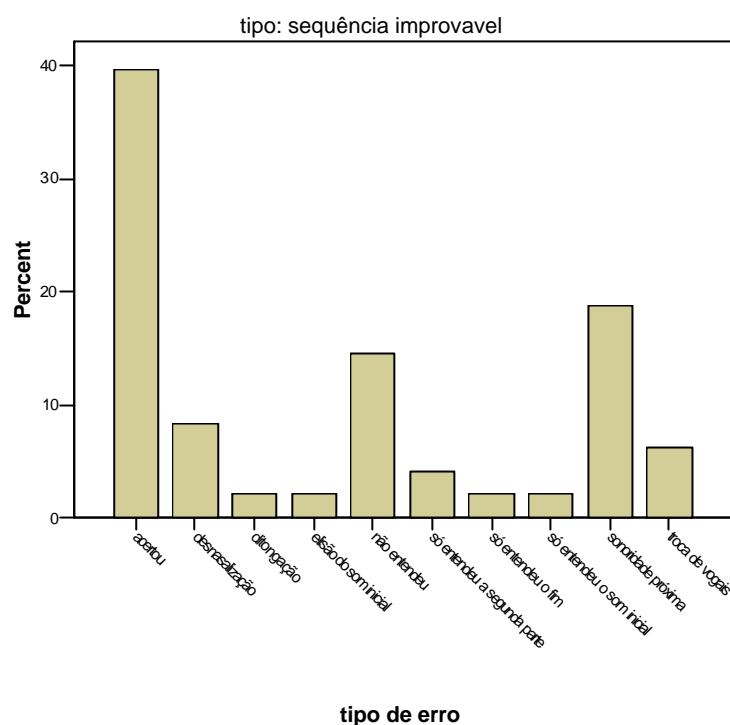


Figura 12- Ocorrência de erros na sequência improvável

As sequências improváveis apresentam a percentagem mais baixa de acertos, devido à sua dificuldade de percepção, uma vez que são improváveis na Língua Portuguesa. Assim, apenas 40% se referem à percentagem de acerto. Os casos de sonoridade próxima atingem cerca de 20%, o que, muito provavelmente, resulta da dificuldade que este tipo de sequências. Estes números indicam que o ouvinte “encontrou” uma sequência de sonoridade semelhante que, sendo comum no seu vocabulário e, por isso, mais familiar, substitui a improvável.

Os exemplos de palavras que não foram de todo entendidas alcançam também uma percentagem de quase 15%, o que comprova a dificuldade na percepção destas sequências. A sua inclusão na nossa lista tornou-se, no entanto, necessária para contemplar todos os difones.

Não podemos deixar de referir também que, neste tipo de sequências, o tipo de erro abrange quase todas as categorias, à exceção do acrescento de som inicial.

Ao realizarmos os gráficos em função da avaliação da qualidade das vozes, obtivemos os resultados que seguidamente apresentamos. Cada um dos gráficos mostra, separadamente, os resultados obtidos para cada um dos estímulos.

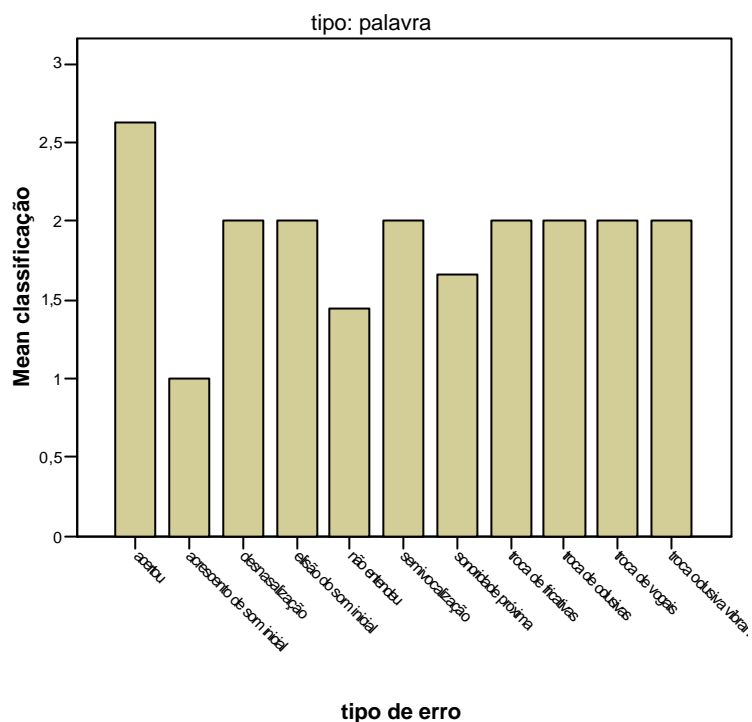


Figura 13- Avaliação da qualidade da palavra

Na figura 13, note-se a média acima de 2.5 para as palavras em que houve acerto. Um facto curioso é que, mesmo quando os ouvintes não entendem as palavras totalmente, em média, atribuem-lhes a classificação de dois, considerando-as, por isso, realizações satisfatórias.

Se tivermos em conta a classificação mais baixa – de 1 ponto –, que caracteriza as realizações de má qualidade, ela é atribuída a casos de acrescento do som inicial e não, como seria de prever, às palavras que os ouvintes não entenderam, tendo estas obtido uma média de quase 1.5.

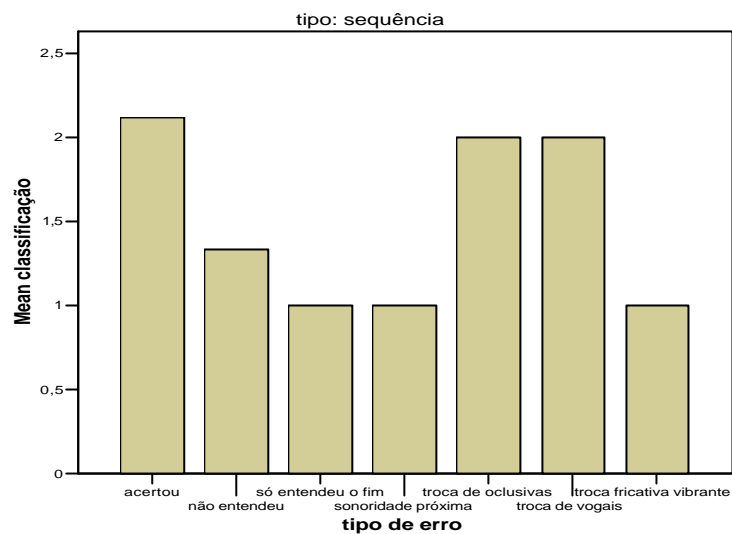


Figura 14- Avaliação da qualidade das sequências.

No que diz respeito às sequências, a classificação dois é atribuída simultaneamente a três tipos de categorias: os acertos, a troca de oclusivas e a troca de vogais.

As sequências que os ouvintes não entenderam obtiveram uma classificação superior às restantes categorias.

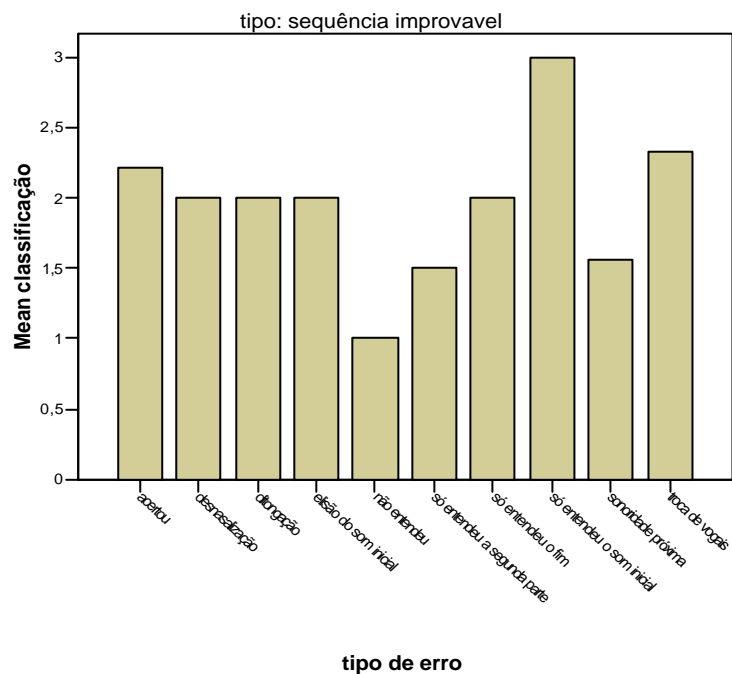


Figura 15 - Avaliação da qualidade das sequências improváveis.

Uma característica surpreendente em relação à qualidade das sequências improváveis é a classificação atribuída nos diferentes casos. Apenas três categorias - sequências não entendidas, onde só foram entendidas a segunda parte e as trocas por sequências de sonoridade próxima - foram classificadas com valores abaixo do nível satisfatório.

Note-se também que o único estímulo que apresenta a avaliação máxima, perto dos três pontos, refere-se às sequências onde só foi entendido o som inicial. Esta classificação deve-se, provavelmente, ao facto dos ouvintes considerarem que este som inicial se tratava do único estímulo a ser realizado.

As realizações das sequências improváveis obtiveram a melhor avaliação, tendo uma classificação global acima dos 2 pontos, atingindo, por isso, o nível satisfatório.

5.1.5 Resultados para a voz com pronúncia da variante do Porto

Nestes segundos testes, é notório que, apesar de termos trocado a ordem de execução dos mesmos, os ouvintes já estavam, de certo modo, familiarizados com a voz sintética, o que gerou um menor número de enganos e incorrecções. Os testes realizados para a voz com a pronúncia do Porto foram sujeitos exactamente aos mesmos processos de análise, sendo os gráficos construídos, tendo em conta as mesmas variáveis.

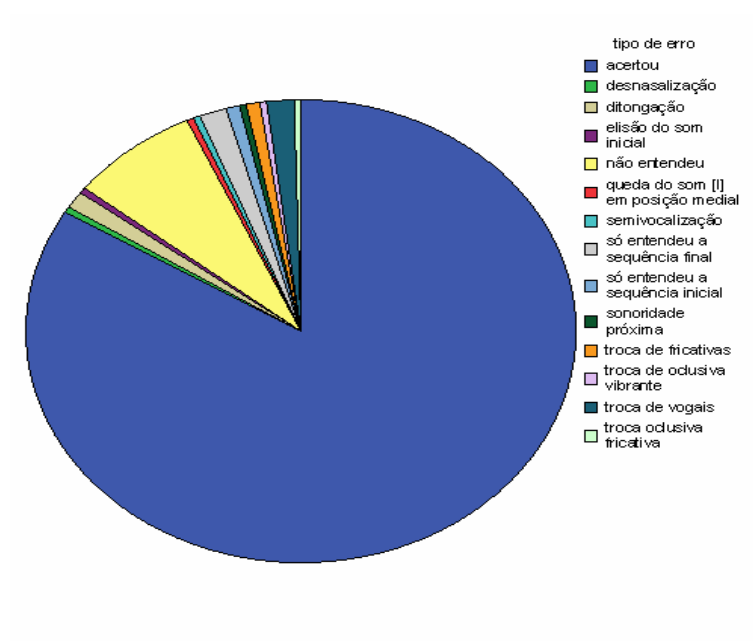


Figura 16 – Resultados globais para a voz com pronúncia do Porto.

Numa primeira abordagem, não podemos deixar de realçar a elevada percentagem de acerto dos ouvintes (a azul), no teste de percepção na globalidade. Apenas uma percentagem pouco significativa corresponde à ocorrência dos diferentes tipos de erros.

Para uma análise mais pormenorizada, apresentamos os valores, também na tabela 3.

Tabela 3- Resultados globais e percentuais para a voz com pronúncia do Porto

		Tipo de estímulo					
		palavra		sequência		sequência improvável	
		Casos	% palavras	Casos	% sequência	Casos	% seq improvável
	acertou	154	89,0%	21	84,0%	34	65,4%
tipo de erro	desnasalização	0	,0%	0	,0%	1	1,9%
	ditongação	0	,0%	1	4,0%	2	3,8%
	elisão do som inicial	0	,0%	0	,0%	1	1,9%
	não entendeu	11	6,4%	1	4,0%	7	13,5%
	queda do som [l] em posição medial	1	,6%	0	,0%	0	,0%
	semivocalização	1	,6%	0	,0%	0	,0%
	só entendeu a sequência final	1	,6%	1	4,0%	2	3,8%
	só entendeu a sequência inicial	0	,0%	1	4,0%	1	1,9%
	sonoridade próxima	1	,6%	0	,0%	0	,0%
	troca de fricativas	2	1,2%	0	,0%	0	,0%
	troca de oclusiva/vibrante	0	,0%	0	,0%	1	1,9%
	troca de vogais	2	1,2%	0	,0%	2	3,8%
	troca oclusiva/fricativa	0	,0%	0	,0%	1	1,9%

A tabela 3, à semelhança da tabela 2, apresenta a informação sobre os tipos de estímulos, as categorias de erro, bem como as suas percentagens.

A percentagem de acerto na palavra foi de 89%, sendo os restantes 11% representantes dos tipos de erros ocorridos. As sequências obtiveram um grau de acerto de

cerca de 84% e 16% de erros. Nas sequências improváveis, devido ao seu maior grau de dificuldade, o acerto é menor, perto dos 65,4%, e 34,6% de erros.

Os tipos de erros, para cada tipo de estímulo, serão descritos em cada um dos gráficos que se seguem:

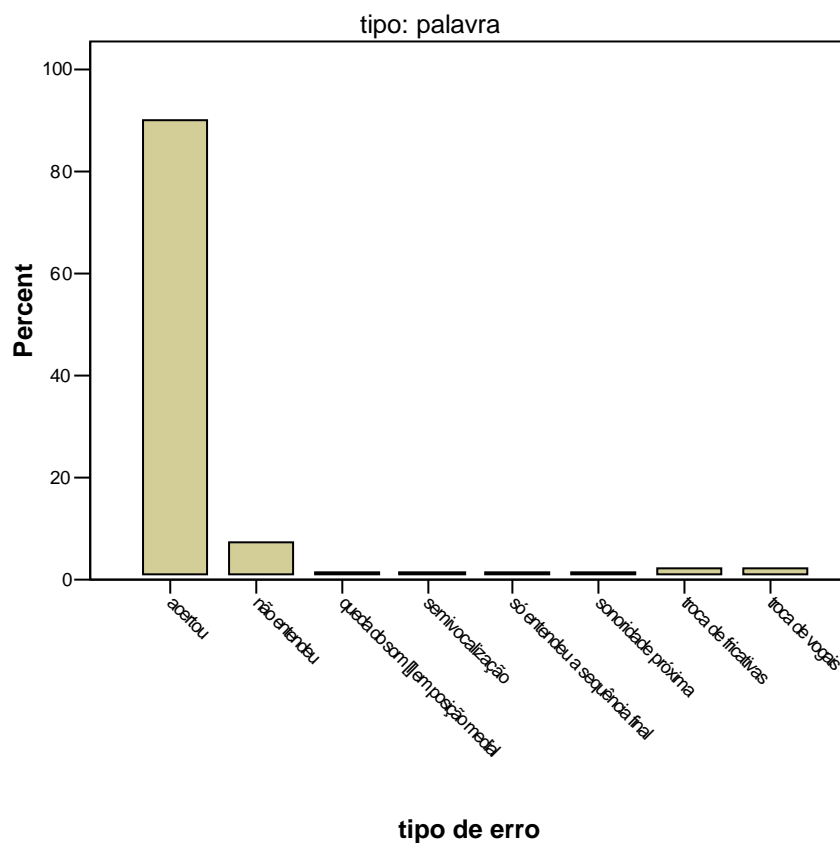


Figura 17 - Realização de tipos de erros nas palavras na voz com pronúncia do Porto

Ao analisarmos a figura 17 e os erros ocorridos, concluímos que apenas as palavras não entendidas atingem uma percentagem significativa. Os restantes erros, pelo seu número reduzido de ocorrência, não merecem uma análise detalhada.

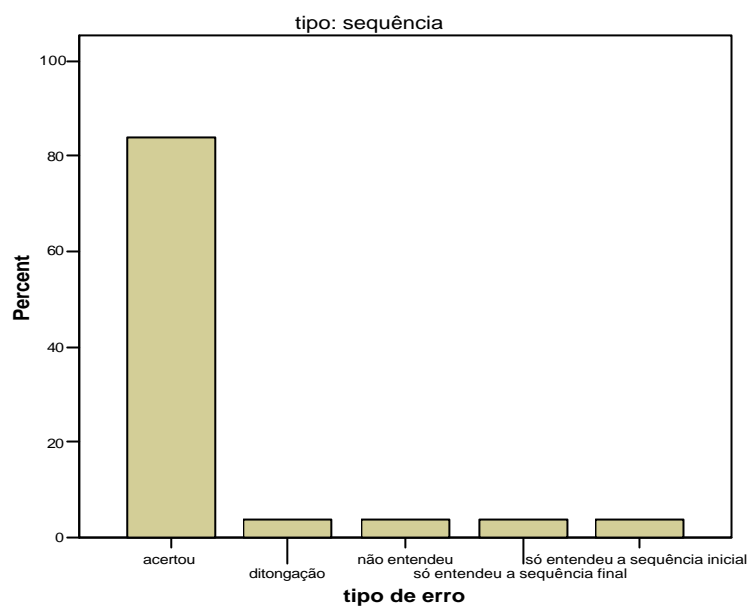


Figura 18 – Realização de tipos de erros nas sequências na voz com pronúncia do Porto

Ao nível da sequência, os erros espalham-se por menos categorias, apenas ocorrendo cinco tipos de erros, em vez dos onze encontrados no estímulo - palavra.

Os erros encontrados - ditongação, não entendeu, só entendeu a sequência final e só entendeu a sequência inicial - obtiveram a mesma percentagem, sendo os três últimos relativos à não percepção total ou parcial do estímulo.

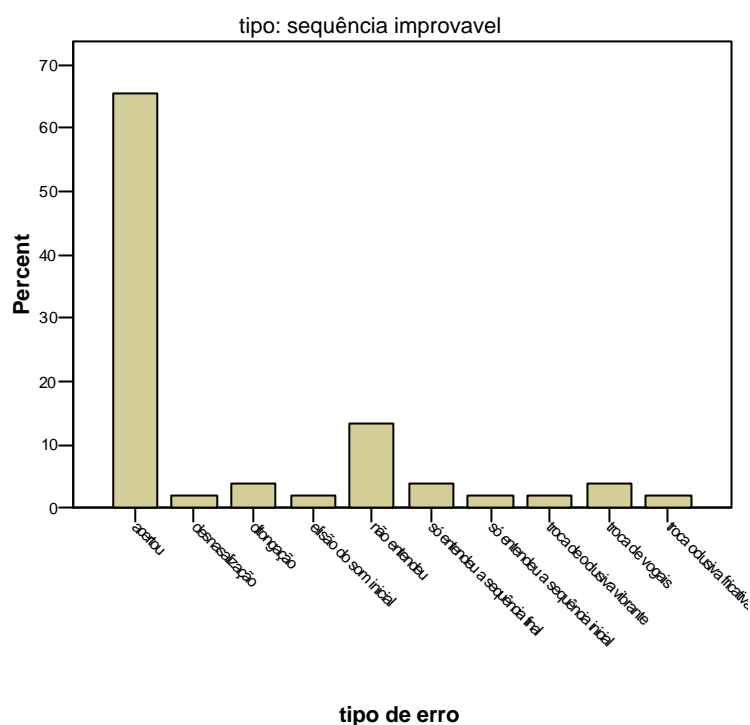


Figura 19 – Realização de tipos de erros nas sequências improváveis na voz com pronúncia do Porto

Nas sequências improváveis, a percentagem de acerto é menor, o que evidencia as dificuldades na percepção das mesmas.

A ocorrência de erros pode ser dividida em três grupos: com cerca de 10% encontram-se as sequências não entendidas, i. e., sem qualquer resposta por parte do ouvinte; com uma percentagem menos significativa, surgem os erros de ditongação, a troca de vogais e a percepção parcial da sequência, neste caso da parte final; com um valor que não podemos considerar estatisticamente significativo, ocorrem a desnasalização, a troca de consoante oclusiva por consoante vibrante, a troca de consoante oclusiva por consoante fricativa e a percepção da parte inicial da sequência.

Tendo-se constatado que, nos testes de qualidade relativos à voz standard, a análise individual dos vários tipos de estímulos não conduziu a resultados muito diferenciados, optámos, no caso da voz do Porto, por fazer uma análise conjunta dos três estímulos.

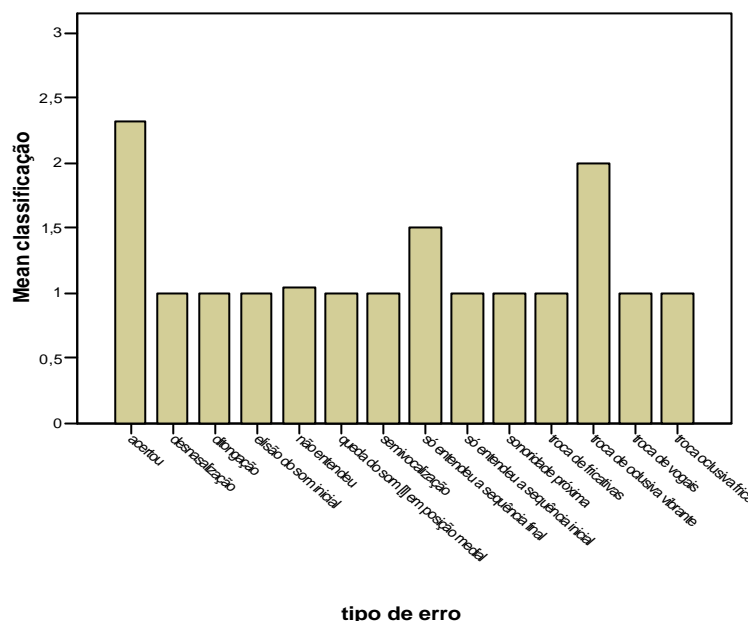


Figura 20 – Avaliação da qualidade da voz do Porto, nos três estímulos (pal., seq. seq. imp.)

Ao analisarmos a classificação atribuída às palavras, às sequências e às sequências improváveis, em termos de qualidade da voz, verificámos que, em casos de acerto, a média de qualidade é de aproximadamente 2,5.

Apesar da troca de consoante oclusiva por consoante vibrante, os ouvintes continuam a considerar a qualidade da voz satisfatória (média de 2 pontos); as sequências em que apenas a parte final foi entendida foram avaliadas com 1,5, sendo as restantes realizações, em que ocorreram outros tipos de erros, avaliadas como de má qualidade.

5.1.5 Discussão

As vozes apresentam um nível similar, obtendo ambas em média uma percentagem de acerto superior a 70%, com vantagem para a voz do Porto.

A sua inteligibilidade é satisfatória, uma vez que os ouvintes conseguiram reproduzir grande parte dos estímulos.

Ao nível da qualidade, os resultados foram os esperados, uma vez que se trata de uma voz sintética, sem qualquer informação prosódica, o que a torna estranha e de difícil

percepção para o ouvido humano. No entanto, a qualidade, no geral, obteve uma média satisfatória.

Este teste não era de todo fácil, uma vez que o *corpus* era constituído por palavras e sequências de palavras do Português Europeu, nem sempre muito comuns, sendo a realização de algumas delas pouco provável ou de todo improvável. Apesar de termos disso consciência, essas mesmas sequências, pelas razões já aludidas, não poderiam ser ignoradas na realização desta pesquisa.

Concluindo: as vozes são satisfatórias, não muito distantes entre si, em termos de inteligibilidade/qualidade, sendo que nenhuma das variantes está em posição de destaque.

5.2 Teste perceptual para avaliação da qualidade da voz

Para a avaliação da inteligibilidade dos sinais produzidos, foi efectuado um teste para a avaliação das vozes. De entre os vários testes possíveis, foi escolhido o mais utilizado: o método de classificação de estímulos em cinco categorias, entre o mau e o excelente, que corresponde a uma pontuação de um a cinco e do qual resulta uma opinião média. Este teste é habitualmente designado por teste Mean Option Score (MOS) (Quackenbush et al., 1998; Rothausen, 1969; citados em Teixeira 2000).

5.2.1- Procedimento

O teste MOS foi realizado recorrendo a um programa de computador, desenvolvido no âmbito do trabalho de Doutoramento de um dos orientadores (Teixeira, 2000). A interface com o utilizador permite aos ouvintes atribuir a classificação utilizando o rato. Por outro lado, estes podem também ouvir o estímulo quantas vezes desejarem. Só depois de classificarem cada estímulo é que o computador passa ao estímulo seguinte.

A pergunta realizada aos ouvintes foi: “A voz é do Porto?”

Na classificação dos estímulos, os ouvintes utilizaram a seguinte escala:

1. Não é de certeza do Porto
2. Não parece ser do Porto
3. Talvez sim, talvez não, não consigo decidir

4. Talvez seja do Porto
5. É com muita certeza do Porto

No sentido de obter informação acerca da consistência das respostas de cada ouvinte, assim como conseguir um maior número de avaliações, cada estímulo, no nosso caso frase, foi repetido quatro vezes, duas para a voz standard e duas para a voz do Porto. Na construção do teste, os estímulos foram distribuídos aleatoriamente pelo computador. Todos os ouvintes foram confrontados com as mesmas frases, apresentadas pela mesma ordem. O teste foi realizado individualmente, numa sala com ruído baixo a moderado, sendo os estímulos apresentados a um ouvinte de cada vez, com utilização de auscultadores.

5.2.2- Estímulos

Optámos por escolher pequenas frases que foram retiradas do jornal *O Público*, nos dias 31 de Janeiro e 9 de Fevereiro de 2005 e que, depois de normalizadas, isto é, com os números, as siglas e as abreviaturas tratados, constituem o nosso *corpus*. A tabela que seguidamente apresentamos inclui as frases utilizadas.

Tabela 4 – Frases seleccionadas para o teste

líder palestiano desaparece aos setenta e cinco anos
patrões acusam partidos de actuar por interesse próprio
quase metade das famílias portuguesas habita em casas frias
começou o julgamento do processo mais mediático dos últimos anos
futebol clube do porto foi hoje agraciado pelo governo com a medalha de mérito turístico
este é um espectáculo de sapateado pouco ou nada convencional
sociedade de informação é fundamental para combater morosidade
estamos com a selecção nacional em todos os desafios
lisboa antiga vibrou com o desfile carnavalesco
bloco diz que a igreja não deve tomar posições políticas
camião descontrolado mata trinta e oito pessoas em angola
a menina dos beijos
o dever de votar

Devemos referir que, ao escolher as frases, houve a preocupação de seleccionar algumas, como é o caso da frase “o dever de votar”, onde os fenómenos característicos da variedade do Porto tivessem maior probabilidade de ocorrência. O nosso objectivo era auxiliar o ouvinte na sua decisão, permitindo-lhe reconhecer a voz como sendo do Porto e avaliando-a, por isso, com o máximo da pontuação.

Todo o texto foi reescrito em letras minúsculas, exigência necessária para a execução do programa de conversão grafema-fone e o tratamento da informação.

Após a normalização, procedemos à transcrição das frases recolhidas, usando um sistema automático de conversão grafema-fone, baseado em regras manuais que se encontra em desenvolvimento no Centro de Línguas e Culturas da Universidade de Aveiro (Oliveira *et al*, 2004).

Em seguida, efectuámos manualmente a correcção/verificação das anotações geradas pelo sistema automático, de forma a colmatar as lacunas existentes. Por exemplo para a frase:

“futebol clube do porto foi hoje agraciado pelo governo com a medalha de mérito turístico”

obtivemos:

**#fut@bOlklub@#du#portu#foj#oZ@#6gr6siadu#pelu#g
uvernuko~#6#m@daL6d@#mEritu#turiStiku#**

De seguida, procedemos ao tratamento de algumas pausas, uma vez que, como referido, a nossa voz não tem implementados os módulos referentes à prosódia.

**#fut@bOlklub@duportufoj oZ@ 6gr6siadu pelu guvern
u#ko~ 6 m@daL6d@ mEritu turiStiku#**

O mesmo procedimento foi aplicado a todas as frases e encontra-se em anexo (anexo 4).

Finalmente, adaptámos manualmente a transcrição standard para a pronúncia do Porto, tendo sido obtidas, tomando ainda como exemplo a frase acima transcrita, as duas versões que se seguem:

pronúncia standard -#f u t @ b O l k l u b @ d u p o r t u f o j o Z @ 6 g r 6 s i a d u
p e l u g u v e r n u # k o ~ 6 m @ d a L 6 d @ m E r i t u t u r i S t i k u #

pronúncia do Porto- #f u t @ b O l @ k l u b @ d u p w 6 r t u f o j @ o j Z 6 g r 6 s i a
d u p e l u g u b e r n u k o ~ 6 m @ d a j L 6 d @ m E r i t u t u r i S t i k u #

As transcrições obtidas para cada uma das frases foram sintetizadas, utilizando ambas as vozes. Desta forma, foi possível obter quatro estímulos para cada frase: a voz standard (criada a partir das gravações realizadas pelo falante da variedade normativa) com as regras standard, a voz standard com as regras da pronúncia do Porto, a voz do Porto (construída com base nas realizações de um informante natural do Porto) com as regras standard e a voz do Porto com as regras da pronúncia do Porto.

5.2.3- Os ouvintes

Os ouvintes foram escolhidos arbitrariamente, não sendo exigida nenhuma característica específica, garantindo apenas que estes não conheçam as vozes em análise, nem o conteúdo dos testes. Para assegurar o cumprimento deste requisito, os ouvintes que participaram neste teste foram distintos dos que realizaram os testes anteriormente descritos. Nestes testes, participaram 6 ouvintes, três do sexo masculino e três do sexo feminino, todos tendo o Português como língua materna, naturais e residentes, na sua maioria, na zona litoral norte do país. As idades variam entre os 22 e os 28 anos. Os níveis de escolaridade cobrem um intervalo situado entre o ensino secundário e a licenciatura, inclusive. Todos os ouvintes têm contacto com o Falar do Porto, apresentando, também eles, algumas características desse Falar. No entanto, não são sensíveis à maior parte das variantes fonéticas desse falar, reconhecendo apenas dois tipos de fenómenos fonéticos: a clássica troca e realização de [v] como [b] e a produção do som [o~] em palavras como *pão*, *são*, *cão*, pronunciadas como [po~], [so~] e [ko~]. Esta realização é entendida como

do Porto e conotada de forma pejorativa. Apesar desse pré-conceito e preconceito, e de acordo com o nosso estudo, este fenómeno não é realizado pela maioria dos falantes do Porto, sendo usado por um número muito reduzido de pessoas na sua maioria iletradas e de idade avançada, ou de forma caricatural.

Tabela 5 – Os ouvintes participantes no teste

Nome	Sexo	Idade	Escolaridade	Residência
AF	Feminino	23	universitária	Porto
RS	Masculino	28	Secundário	Sta. Maria da Feira
SF	Feminino	28	licenciada	Sta. Maria da Feira
JF	Masculino	24	licenciado	Aveiro
EO	Feminino	22	universitária	Coimbra
ZR	Masculino	26	universitário	Viseu

Para aferir a consistência das respostas de cada ouvinte, foi calculada a percentagem das decisões iguais entre as duas repetições da totalidade dos estímulos (coluna % igual).

Como uma pequena diferença é muitas vezes insignificante, calculámos também as vezes em que as respostas diferiam no máximo 1 ponto.

Tabela 6- Consistência das respostas de cada ouvinte entre duas repetições.

Ouvinte	% igual	% similar (dif <=1)
SF	44.2	65.4 !
JF	55.8	80.8
EO	44.2	71.2
ZR	30.8 !	65.4 !
RS	44.2	82.7
AF	48.1	86.5
Média	45.6	73.6

Foi aplicado o mesmo processo para comparar as respostas dos ouvintes, apresentando-se, na tabela 7, a percentagem de vezes em que os ouvintes atribuíram a mesma avaliação.

Tabela 7 – Consistência das respostas entre os ouvintes nas duas repetições.

	SF	JF	EO	ZR	RS	AF
SF	---	35.5769	32.6923	20.1923	38.4615	42.3077
JF		----	25.0000	23.0769	43.2692	34.6154
EO			---	26.9231	25.9615	23.0769
ZR				---	25.9615	20.1923
RS					---	36.5385
AF						---

Na tabela seguinte (tabela 8), encontram-se os resultados em que a resposta diferiu apenas uma unidade.

Tabela 8 - Correlação entre as classificações dos vários ouvintes, com diferença < ou = a 1

	SF	JF	EO	ZR	RS	AF
SF	---	66.3462	62.5000	54.8077	63.4615	74.0385
JF		---	74.0385	60.5769	77.8846	69.2308
EO			---	60.5769	68.2692	61.5385
ZR				---	66.3462	63.4615
RS					---	70.1923
AF						---

Estes resultados permitem-nos comparar as classificações atribuídas pelos diferentes ouvintes, quando a sua resposta diverge apenas num ponto.

Os resultados de um dos ouvintes (ZR) revelaram uma grande inconsistência entre as duas repetições, facto que nos leva a não o considerar para o processamento.

5.2.4- Os resultados

Embora reconheçamos que os resultados do teste ganhariam com um maior número de ouvintes/avaliadores, o limite temporal que nos é imposto por este trabalho, apenas nos permitiu aplicar o teste a seis ouvintes, sendo apenas analisados os resultados de cinco deles, pela razão acima apresentada.

Depois dos dados recolhidos e tratados, pareceu-nos pertinente avaliá-los em função de diferentes variáveis que nos permitissem validar, ou não, a qualidade das nossas vozes, de acordo com o objectivo da criação de uma voz do Porto inteligível .

Numa primeira análise, interessa-nos saber a média atribuída a cada voz, com e sem conversão grafema-fone, adequada à variante em estudo.

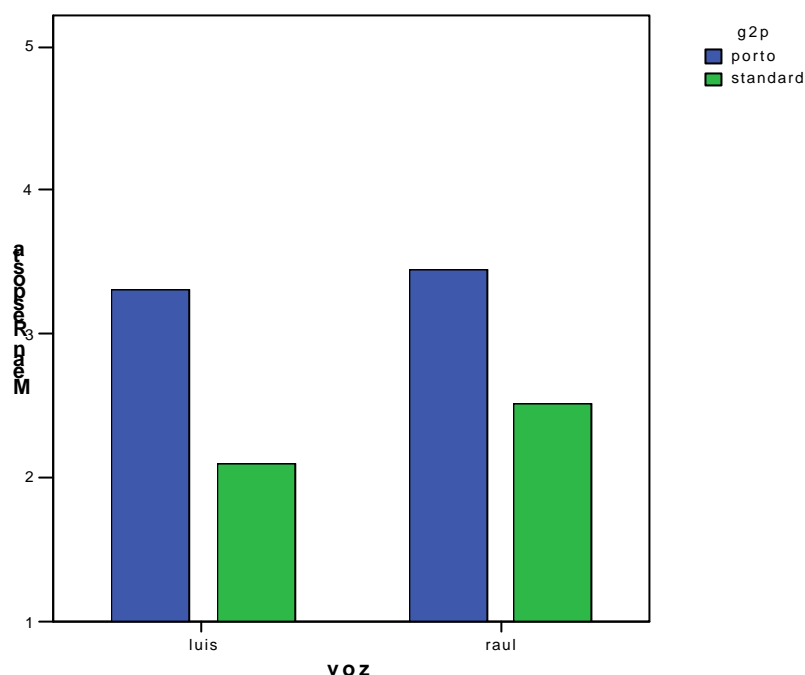


Figura 21 – Média de aceitação das vozes com e sem regras de conversão grafema-fone

Nesta figura 21, encontramos do lado esquerdo a voz do Luís e do lado direito a voz do Raúl. As barras azuis representam as vozes onde foram aplicadas as regras passíveis de caracterizar a voz do Porto. As verdes apresentam as vozes com as regras entendidas como características da variedade normativa.

Em termos médios, as vozes obtêm uma melhor classificação quando são aplicadas as regras do Falar do Porto. Ao aplicarmos as regras de standard, os ouvintes preferem, apesar disso, a voz do Raúl, como podemos observar pelos resultados apresentados na figura.

Se analisarmos as vozes, tendo em conta a variação das respostas (dispersão), usando um intervalo de confiança a 95% para a média, obtemos a figura 22, abaixo apresentada.

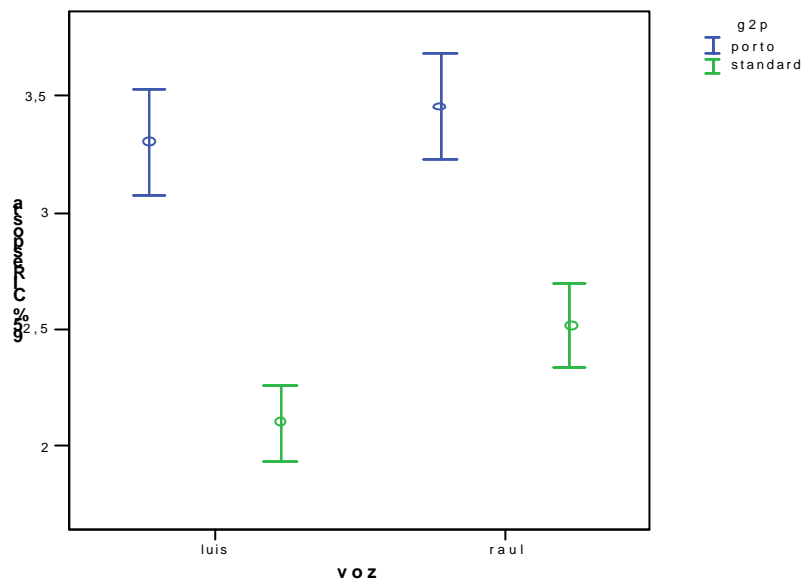


Figura 22- Média de aceitação das vozes com e sem regras de conversão grafema-fone com um intervalo de confiança de 95%

A diferença entre a voz do Raúl e a do Luís, com as regras de conversão standard, é significativa. Os seus intervalos de confiança não se sobrepõem. Também é expressiva a diferença entre as vozes com e sem a utilização das regras de conversão grafema-fone do Porto. Os ouvintes continuam a preferir as vozes a que foram aplicadas as regras que traduzem fenómenos característicos do Falar do Porto.

Numa segunda análise, considerámos pertinente avaliar os resultados da avaliação de cada ouvinte individualmente.

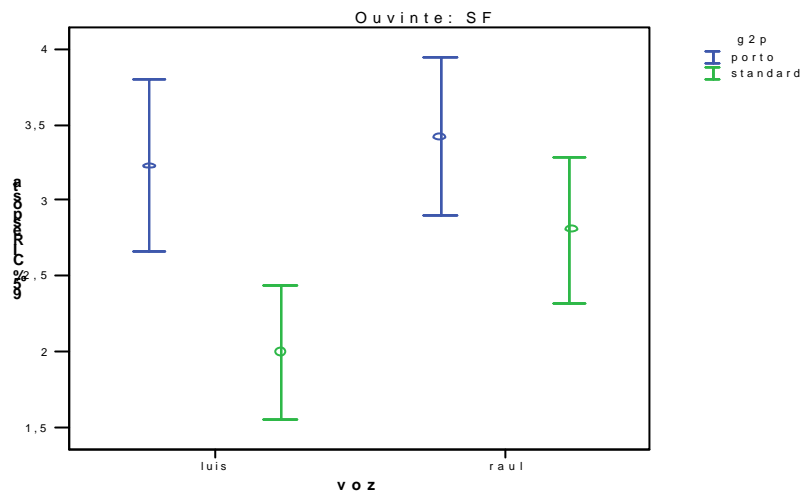


Figura 23 – Resultados da aceitação das vozes para a ouvinte SF

A primeira ouvinte (SF), ao avaliar as vozes com a conversão grafema-fone standard, prefere a voz do Raúl, em detrimento da voz do Luís, não chegando a preferência a ser significativa, devido à grande amplitude dos intervalos. Tratando-se da voz com a conversão grafema-fone do Porto, a avaliação das vozes é muito semelhante, tendo, no entanto, a voz do Raúl obtido uma pontuação média ligeiramente mais alta.

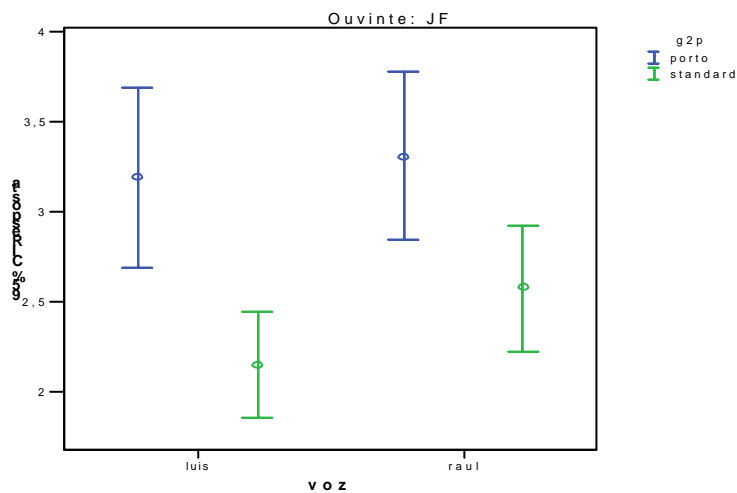


Figura 24- Resultados da aceitação das vozes para a ouvinte JF

Em relação ao segundo ouvinte (JF), os resultados são semelhantes. Este, ao avaliar as vozes sem as regras do Porto, prefere a voz do Raúl, sendo, no entanto, a diferença de pontuação pouco expressiva estatisticamente.

Se adicionarmos às vozes as regras do Porto, embora com uma diferença mínima, a voz do Raúl é mais aceite como sendo do Porto.

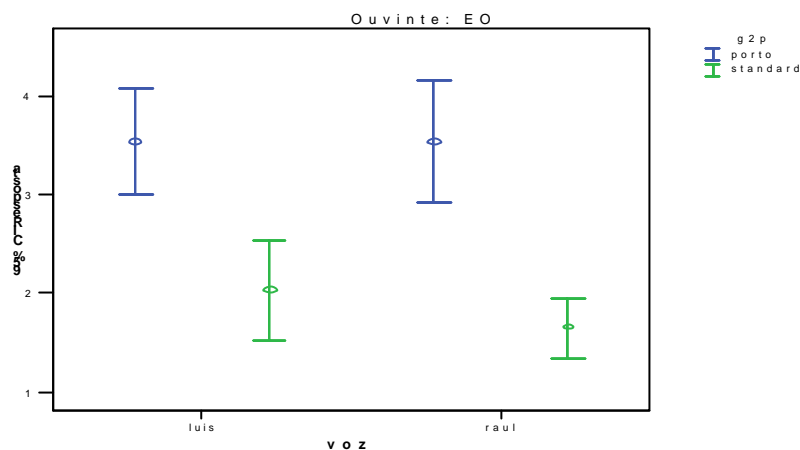


Figura 25 - Resultados da aceitação das vozes para a ouvinte EO.

Os resultados da ouvinte EO são um pouco diferentes. A ouvinte, ao avaliar as vozes sem as regras do Porto, prefere a voz do Luís, muito provavelmente porque este possui um sinal mais forte e audível, o que a torna mais agradável ao ouvido.

Se adicionarmos, às vozes em análise, as regras do Porto, a ouvinte classifica-as, em média, com a mesma pontuação.

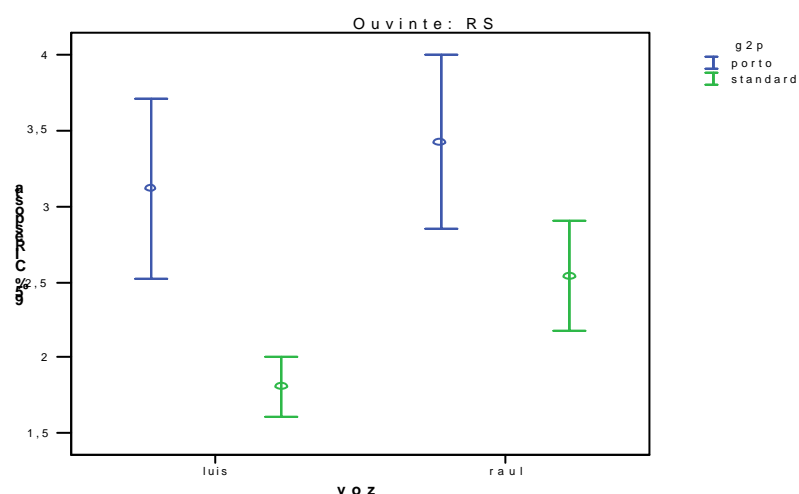


Figura 26 - Resultados da aceitação das vozes para a ouvinte RS.

Sem as regras do Porto, este ouvinte (RS) prefere, sem dúvida, a voz do Raúl, sendo a diferença bastante significativa estatisticamente, visto não haver sobreposição dos intervalos. Na avaliação, com as regras do Porto, a média de ambas as vozes apresenta-se muito semelhante. Mesmo assim, e apesar disso, a voz do Raúl é preferida, obtendo uma melhor pontuação.

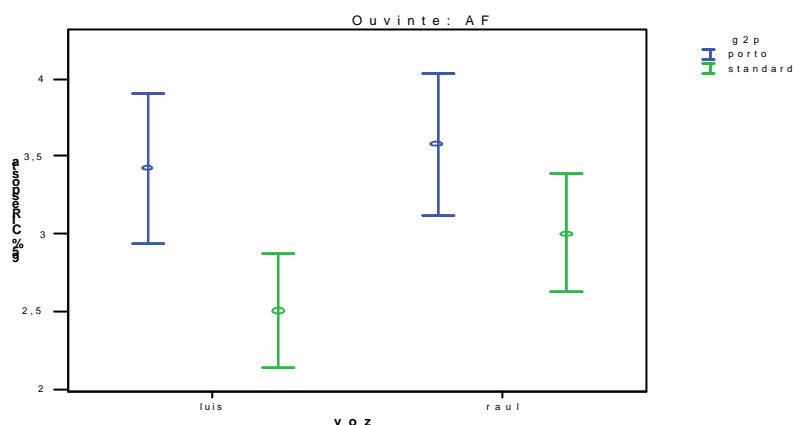


Figura 27 - Resultados da aceitação das vozes para a ouvinte AF

Os resultados da quinta ouvinte (AF) são semelhantes aos dos ouvintes SF, JF e RS. A voz do Raúl é preferida. No entanto, essa diferença é mais notória sem as regras do Porto.

Numa parte dos casos, a diferença entre as vozes não é significativa, mas a relação de ordem é similar aos resultados gerais obtidos: a voz do Raúl é avaliada com melhor pontuação, mas esta diferença diminui, em termos gerais, aquando da adição das regras de conversão grafema-fone relativas à voz do Porto.

Numa terceira análise, exemplificamos e apresentamos algumas frases, cujos resultados são mais pertinentes para este estudo. As figuras representam a avaliação média atribuída por todos os ouvintes a cada uma das frases em análise.

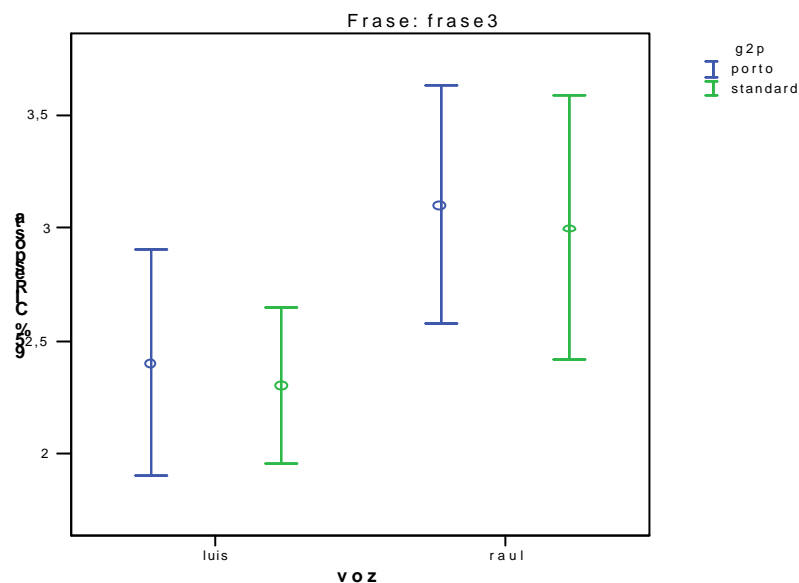


Figura 28 - Resultados da aceitação das vozes para a frase 3

A figura 28 mostra que, na frase três - “quase metade das famílias portuguesas habita em casas frias” -, quando aplicadas as regras de conversão grafema-fone relativas à voz do Porto, melhora a qualidade das duas vozes. Manifesta-se, ainda assim, uma maior aceitação da voz do Raúl, mesmo aplicando as regras do Porto à voz do Luís.

A voz parece ser determinante para distinguir as produções com pronúncia standard das produções com pronúncia da variante do Porto. Para efeitos de síntese, parece ser importante que a voz gravada contenha já algumas características da variante que se pretende sintetizar.

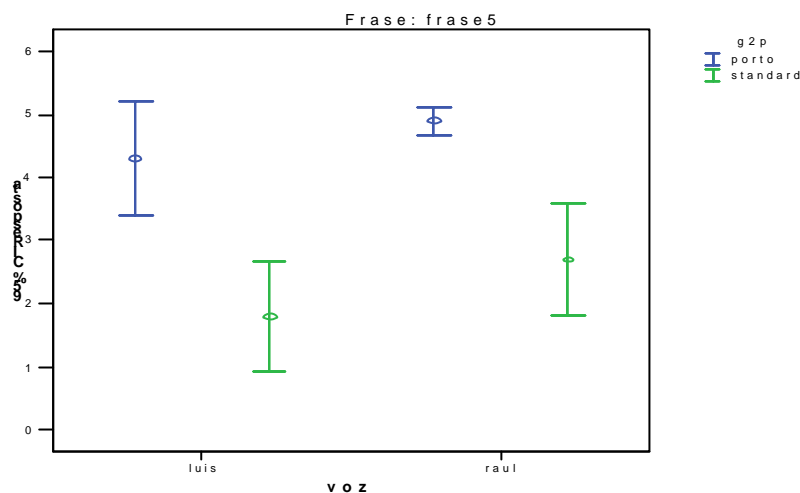


Figura 29 - Resultados da aceitação das vozes para a frase 5

A figura 29, remetendo para a frase 5 - “futebol clube do porto foi hoje agraciado pelo governo com a medalha de mérito turístico”-, demonstra que a junção dos dois factores, a voz e as regras de conversão grafema-fone da variante do Porto, permite aos ouvintes classificar, com poucas dúvidas, a frase como sendo realizada por um falante do Porto.

Apesar da voz do Luís, com a aplicação das regras de conversão grafema-fone relativas à variante do Porto, ter uma boa classificação, o intervalo de confiança para a voz do Raúl é de reduzida amplitude, resultando numa média próxima de 5, a pontuação máxima. Quando confrontados com esta frase, a resposta da quase totalidade dos ouvintes foi a seguinte: a voz “é com muita certeza do Porto”.

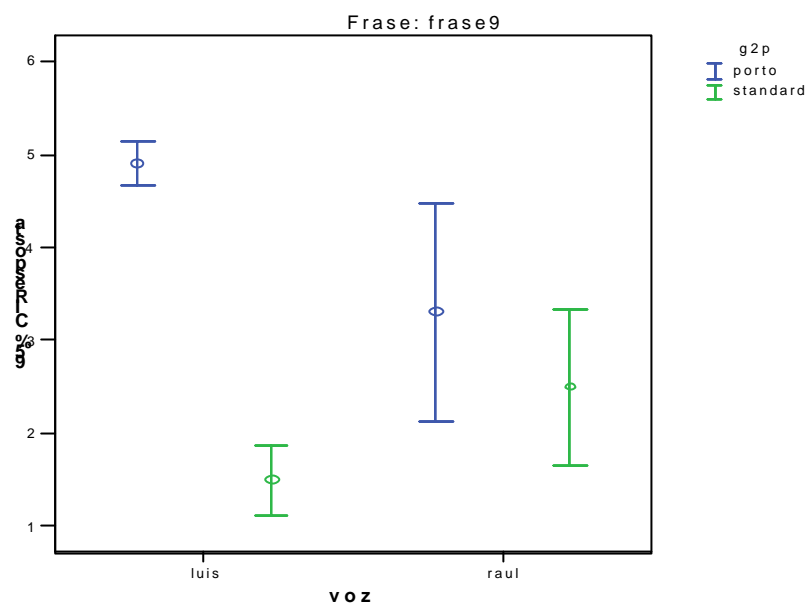


Figura 30 - Resultados da aceitação das vozes para a frase 9.

A nona frase - “lisboa antiga vibrou com o desfile carnavalesco” - apresenta um resultado completamente diferente de todos os restantes. A voz do Luís, com as regras de conversão grafema-fone do Porto, é classificada como sendo “com muita certeza do Porto”.

A voz do Raúl, com e sem as regras, obtém resultados pouco diferenciados.

Ao compararmos as vozes dos dois informantes, com a aplicação das regras de conversão grafema-fone relativas à voz do Porto, encontramos uma diferença significativa, uma vez que os intervalos de confiança quase que não se sobrepõem.

Porém, a voz do Luís, com regras standard, continua a obter uma classificação muito baixa, portanto muito semelhante à das outras frases.

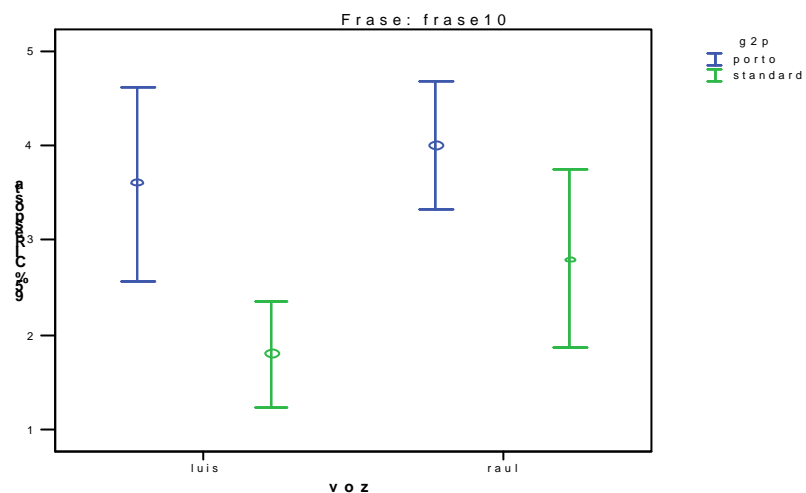


Figura 31- Resultados da aceitação das vozes para a frase 10.

Na frase 10 - “bloco diz que a igreja não deve tomar posições políticas” -, confirma-se a preferência pela voz do Raúl.

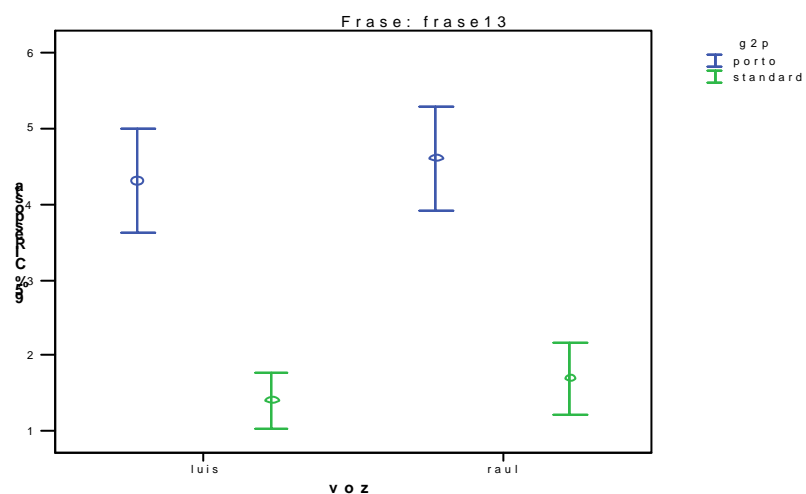


Figura 32- Resultados da aceitação das vozes para a frase 13

Esta última frase - “o dever de votar”-, representada na figura 32, realça a importância das regras e da sua inclusão na criação das vozes.

A diferença nas classificações é, de facto, notória, principalmente quando se trata de frases de tamanho reduzido.

Numa quarta análise, comparámos os resultados, tendo em conta a ocorrência ou não de fenómenos fonéticos característicos do Falar do Porto.

Para cada frase, foram determinados os fenómenos modificadores de pronúncia entre o standard e o Porto. De seguida, foram contadas quantas inserções, alterações e elisões ocorreram, utilizando-se a soma desses valores como indicativo do grau de mudanças.

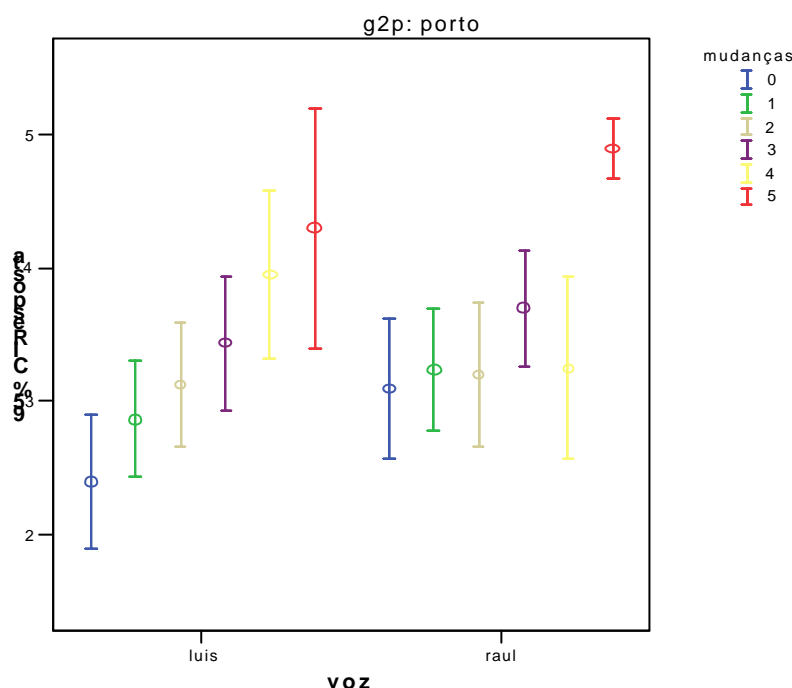


Figura 33- Avaliação das vozes em função do número de mudanças com as regras de conversão grafema-fone do Porto

Analisando a figura 33, tendo em conta a realização das frases quando aplicadas as regras de conversão grafema-fone relativas à voz do Porto, verificamos que as vozes apresentam comportamentos diferentes.

É notório que a classificação obtida pela voz do Luís está intrinsecamente ligada com o aumento da introdução de fenómenos característicos do Porto. Ou seja: quanto mais regras aplicarmos à voz do Luís, mais elevada é a sua classificação. Este aumento é progressivo, como podemos constatar na figura.

A voz do Raúl é mais insensível à inserção dos fenómenos, apresentando-se, de uma forma geral, com um valor médio próximo do 3. Apenas quando lhe adicionamos o

número máximo de regras, e só neste caso, é que o seu valor difere, obtendo uma pontuação muito próxima do 5, a pontuação máxima.

As regras são, sem dúvida, também um factor relevante para a aceitação/rejeição da voz do Porto.

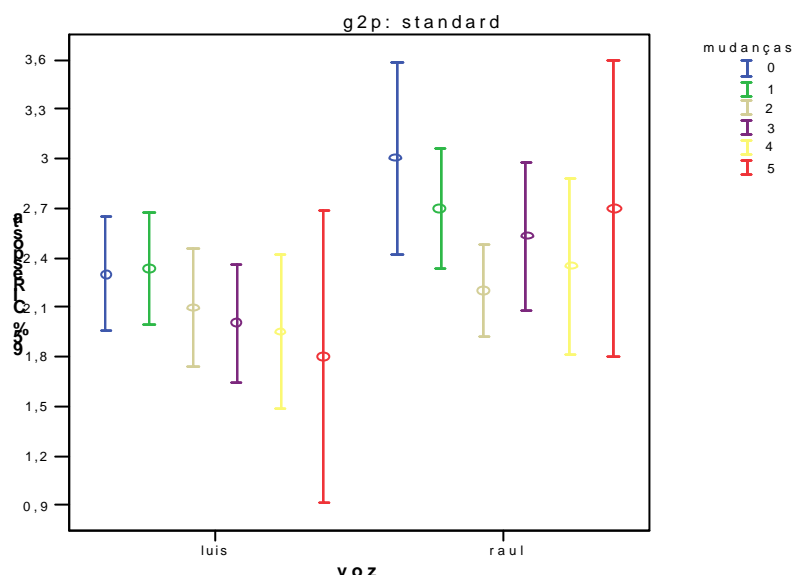


Figura 34 - Avaliação das vozes em função do número de mudanças com as regras de conversão grafema -fone standard

Fazendo uma análise similar, agora com a aplicação das regras de conversão grafema-fone da voz standard, concluímos que, sem as regras características de um falar, a voz propriamente dita, torna-se muito mais importante para a avaliação das realizações produzidas.

A voz do Raúl, agora normalizada, através da aplicação de regras de conversão standard, continua a obter classificações superiores e a ser preferida em detrimento da voz do Luís, daí que possamos concluir que a voz do Raúl continua mais imune em relação às regras.

A classificação da voz do Luís piora, de um modo geral, com a aplicação das regras standard e, sobretudo, em frases em que os fenómenos do Porto tinham uma maior probabilidade de ocorrência.

Concluindo: são necessárias, sem dúvida, as regras do Falar do Porto, mas em casos de não ocorrência de fenómenos característicos da variante em estudo, a voz de um falante da variante é determinante para o resultado. Na criação de uma voz sintética, devemos utilizar, quer a voz de um falante dessa variante, quer as regras desse mesmo falar, de forma a obter um melhor resultado em termos globais, como comprovam os dados que acabámos de apresentar.

6 Conclusões

Esta dissertação descreve a criação e avaliação de uma voz do Porto, através da síntese por concatenação, com base em informação linguística.

Neste último capítulo, apresentamos um resumo do trabalho realizado, as principais conclusões e algumas propostas de trabalho para o futuro.

6.1 *Resumo do trabalho efectuado*

Motivados pela variação linguística do Português Europeu e pela inexistência de vozes sintéticas referentes a essas mesmas variantes, dedicámo-nos, nesta dissertação, à criação de uma voz sintética para o Falar do Porto, que denominámos de Voz do Porto.

Uma parte essencial do trabalho desenvolvido consistiu na criação desta voz, através da síntese por concatenação de difones.

Numa primeira fase, fizemos uma investigação informal, no seio da comunidade de falantes do Porto, de forma a constatar e validar a existência dos fenómenos referenciados pelos autores que, de alguma forma, trataram esta variante. Constatámos que estes fenómenos ainda se mantêm, hoje em dia, continuando a ser motivo de orgulho dos portuenses.

Depois da selecção dos fenómenos a estudar, passámos à construção da lista de fones existentes no PE e na variante do Porto, bem como das possíveis relações entre si (difones). Seguidamente, escolhemos palavras e sequências de palavras onde se realizassem cada uma destas associações.

A lista de palavras, depois de corrigida e transcrita com o alfabeto SAMPA, foi gravada por um falante do Porto e por nós anotada. Numa primeira fase, esta anotação foi

realizada de uma forma automática. Seguiu-se uma correcção manual, de forma a eliminar alguns erros gerados pelo sistema automático. Com base na anotação e no material gravado, utilizando as ferramentas disponibilizadas pelo projecto Festvox, criámos a Voz do Porto.

A construção de uma outra voz, característica da variedade normativa, surgiu da necessidade de avaliar a voz do Porto em comparação com a standard.

De acordo com este propósito, gravámos o mesmo *corpus*, mas desta vez com um informante com características da variedade normativa, repetindo todo o processo da criação da voz.

As vozes obtidas foram, na fase final do trabalho, validadas através de testes de percepção: um de identificação, para a avaliação da inteligibilidade das vozes e identificação dos seus problemas; um outro para testar se a voz era, de facto, percebida como característica da variante do Porto.

6.2 Resultados principais

O resultado principal deste trabalho é a voz criada. Esta poderá ser usada como ferramenta em testes de percepção, no sentido de testar as diferentes formas de realização da mesma palavra.

Outro dado importante é a confirmação, com base nos resultados obtidos, de que a criação de vozes relativas a variantes, como o Falar do Porto, passa, não só pela criação e aplicação de regras de pronunciação adequadas, mas também pelo recurso a falantes dessa variante.

Para além disso, constatámos que, quando a variante em estudo se afasta da realização considerada standard, a inclusão das regras de pronunciação dessa variante permite obter bons resultados. No entanto, em muitas situações, principalmente quando a variante se aproxima da realização considerada standard, é suficiente a utilização de uma voz da variante para a obtenção de bons resultados.

6.3 Trabalho Futuro

Um trabalho desta natureza tem de se apresentar forçosamente lacunar, sobretudo pelo seu objecto de estudo, ficando sempre alguns aspectos por abordar e outros que mereceriam um tratamento mais exaustivo.

Nesse sentido, deixamos algumas sugestões de trabalho futuro que, por falta de tempo, não foram abordadas. Para além de outros aspectos que poderiam ser considerados, destacamos os seguintes:

Realização de testes mais exaustivos

Achamos que a esta pesquisa poderiam ser acrescentados novos dados, se os testes realizados apresentassem um carácter mais exaustivo. De facto, para a continuação deste trabalho é fundamental alargar o número de ouvintes, assim como aumentar os estímulos, de forma a obtermos mais informações e resultados.

Melhoramento da voz

Para um incremento da qualidade da voz, um dos aspectos que consideramos fundamental é o tratamento da informação prosódica destas vozes, de forma a poderem ser utilizadas como ferramentas de trabalho noutros estudos. A inclusão de características prosódicas, permitir-nos-ia obter uma voz com mais qualidade e, por isso, mais próxima do discurso humano.

Criação de novas vozes

Seria também importante criar novas vozes características de outras variantes de PE, de forma a trabalharmos e validarmos a nossa Língua como um todo. Este estudo só é possível se considerarmos todas as variantes do Português Europeu, de forma a termos um *corpus* completo, assim como várias vozes sintéticas passíveis de serem utilizadas como ferramenta em vários estudos linguísticos. Variantes como o falar do Alentejo, do Algarve e de S. Miguel, entre outras, são candidatos óbvios e desejáveis para este tipo de investigação.

Bibliografia

- Alvar, Manuel, (1961), Hacia los conceptos de lengua, dialecto e hablas in *Nueva Revista de Filologia Hispânica*, 15, p. 51-60
- Andrade, Amália, (1994). *Reflexões sobre o “e mudo” em Português Europeu*, Congresso Internacional APL
- Black, Alan W. and Lenzo, Kevin A. (2003) *Building Synthetic Voices*, Language Technologies Institute, Carnegie Mellon University and Cepstral LLC
- Barbosa, J.M, (1994). *Introdução ao estudo da Fonologia e Morfologia do Português*, Coimbra, Livraria Almedina
- Boléo, M. de Paiva, (1961). *O Mapa dos dialectos e falares de Portugal Continental*,
- Boléo, M. de Paiva, (1974). *Estudos de Linguística Portuguesa e Românica*, vol.I, Tomo I, Coimbra
- Carvalho, Pedro M.M.L, (2004). *Determinação Automática de segmentos para síntese de fala por Concatenação*, Dissertação de Doutoramento, Lisboa, Universidade Técnica de Lisboa, Instituto Superior Técnico
- Cintra, L.F.L, (1962). Une frontière lexicale et phonétique dans le domaine linguistique portugaise, in *Boletim de Filologia*, Tomo XX, fasc.1 e 2, p. 31-39
- Cintra, L.F.L, (1972). Nova Proposta de classificação dos dialectos galego-portugueses, in *Boletim de Filologia*, Tomo XXII, fascículo 1e 2, Lisboa
- Cintra, L.F.L, (1983). *Estudos de Dialectologia Portuguesa*, Nova Universidade Linguística, Lisboa, Sá da Costa Edições
- Cunha, Celso e Cintra, L.F.L., (1984). Fonética e Fonologia in *Nova Gramática do Português Contemporâneo*, Lisboa, Sá da Costa edições
- Delgado Martins, M.R., (2002). *Fonética do Português, 30 anos de investigação*, Colecção Universitária, Caminho
- Dijkstra, Jelske, (2004). *A first step towards Frisian TTS*, MA Thesis Taalwetenschap, Speechcommunication and Technology, University of Amsterdam
- Dutoit, Thierry, (1997). *An introduction to Text-to-Speech Synthesis*. Kluwer, Dordrecht

- Faria, Isabel, (1996). *Introdução à Linguística Geral e Portuguesa*, série Linguística, Lisboa, Caminho
- Gonçalves Viana, (1973). *Estudos da Fonética Portuguesa*, Imprensa Nacional, Casa da Moeda, Lisboa
- Klatt, D. (1987). Review of Text-to-speech conversion for English, *Journal of the Acoustical Society of America*, JASA, vol 82(3), p.737-793
- Lemmetty, Sami, (1999). *Review of Speech Synthesis Technology*, Master's Thesis, Helsinki University of Technology, Department of Electrical and Communications Engineering
- Lingaard, R., (1985), *Electronic Synthesis of Speech*, Cambridge, Cambridge University Press
- Machado, José Pedro,(1994). *Estrangeirismos na Língua Portuguesa*, Lisboa, Editorial Notícias
- Mateus, M.H.M. (1982). *Aspectos da Fonologia Portuguesa*, Lisboa, Textos de Linguística, Edição do Instituto Nacional de Investigação Científica, 6
- Mateus, M.H.M. e E. d'Andrade, (2000). *The phonology of portuguese*, London, Oxford University Press
- Mateus, M.H.M. et alli (1983). *Gramática da Língua Portuguesa*, Coimbra, Almedina
- Mateus, M.H.M., (1986). A Língua Portuguesa – unidade e diversidade, in *Actas do 1º Encontro de APL*, Lisboa, p. 145-163
- Mateus, M.H.M., (1990). *Fonética, Fonologia e Morfologia do Português*, Lisboa, Universidade Aberta
- Mosser,Von, *Introduction to Speech Synthesis* (27/01/2000)
- Moutinho, Lurdes de Castro, (1986). *Analyse Sociolinguistique du parler de Porto: Etudes phonétiques et phonologiques*, Thèse pour le Doctorat de L'U.S.H.S, Université des Sciences Humaines de Strasbourg
- Moutinho, Lurdes de Castro, (2000). *Uma Introdução ao Estudo da Fonética e Fonologia do Português*, Plátano Edições
- Moutinho, Lurdes de Castro, (2001). *Falar do Porto com todos os "bês"*, Campo das Letras

- Moutinho, Lurdes de Castro; Zerling, Jean Pierre, (2002/2003). Os ditongos orais em Português – estudo acústico preliminar, in Revista da Universidade de Aveiro, Letras, nº 19/20
- Paiva, S., Moutinho, L., Teixeira, A. (no prelo). *Síntese por Concatenação de variantes regionais- o falar do Porto*, XX Encontro da APL, Gulbenkian.
- Oliveira, C. et al. (no prelo). *Um novo sistema de conversão grafema-fone para o PE baseado em transdutores*, II Congresso Internacional de Fonética e Fonologia, Universidade Federal do Maranhão, 2004.
- Sampson, Rodney, (1999). *Nasal Vowel Evolution in Romance*, New York, Oxford University Press
- Schmidt- Radefeldt, Jurgen,(1997). *Dicionário dos anglicanismos e germanismos na Língua Portuguesa*, Lisboa, Editorial Notícias
- Schroeder, M. (1993). A brief History of Synthetic Speech. *Speech Communication*, vol13, p.231-237
- Segura e Saramago;(2001). *Variedades dialectais portuguesas*, in Caminhos do Português, Exposição Comemorativa do Ano Europeu das Línguas (catálogo), Lisboa, Biblioteca Nacional
- Sproat, Richard, (1998). *Multilingual Text-to-Speech: The Bell Labs Approach*, Kluwer Academic Publishers, United States of America
- Teixeira, A. e Vaz, F. (2000). *A Suite of Tcl/Tk Programs for Perceptual Tests*, Instituto de Engenharia Electrónica e Telemática de Aveiro
- Teixeira, A, (2000) *Síntese Articulatoria das Vogais Nasais do Português Europeu*, Tese de Doutoramento, Universidade de Aveiro
- Teixeira, A et al.(2001) *Acerca das Vogais Nasais do Português Europeu*, separata da Revista da Universidade de Aveiro, Letras, Aveiro
- Vasconcelos, José Leite de, (1985). *Opúsculos*, (Dialectologia), org. por Maria Adelaide Cintra, Lisboa, Imprensa Nacional da Moeda
- Vasconcelos, José Leite de, (1987). *Esquisse d'une Dialectologie portugaise*, Lisboa, Edição do Instituto Nacional de Investigação Científica
- Vazquez Cuesta, P. E Mendes da Luz, M.A., (1971) *Gramática da Língua Portuguesa*, Lisboa, Edições Setenta

Endereços electrónicos

<http://emu.sourceforge.net/>

<http://www.festvox.org>

Anexo 1

Alfabeto fonético para o dialecto padrão do Português Europeu

	AFI	SAM-PA	Exemplo
Vogais orais:			
	i	i	vi [vÍ]
	e	e	vê [vé]
	ɐ	ɛ	pé [p ?]
	a	a	pá [pá]
	ɤ	ɔ	cama [c ?m ?]
	ɐ	@	de [d ?]
	ɔ	O	pó [p ?]
	o	o	avô [?vo]
	u	u	tudo [túdu]
Semivogais			
	j	j	pai [páj]
	w	w	pau [páw]
Vogais nasais			
	i	i~	sim [sÍ]
	ɐ	e~	pente [p ?t ?]
	a~	a~	branco [bra~ku]

	õ	o~	ponte [põtʔ]
	u	u~	atum [ʔtu]
Semivogais			
	j~	j~	põe [põj]
	w~	w~	mão [ma~w~]
Consoantes			
	p	p	pá [pá]
	b	b	boda [bod6]
	t	t	tu [tú]
	d	d	dou [dó]
	k	k	quilo [kilu]
	g	g	gato [gátu]
	f	f	facá [fak6]
	v	v	vê [vé]
	s	s	só [sO]
	z	z	casa [kaz6]
	ʔ	S	chave [ʔvʔ]
	ʔ	Z	já [ʔá]
	l	l	lá [lá]
	l	l~	mal [mál]
	ʔ	L	valha [váʔ6]
	m	m	má [má]
	n	n	nada [nad6]
	ʔ	J	senha [s6ʔ6]
	ʔ	r	caro [káru]
	R	R	carro [káRu]

Anexo 2

Tabela de resultados voz com pronúncia standard

Ouvinte	Nº	estímulo	o que ouviram	classificação	tipo	tipo de erro
O1	793	são-lhe		1	sequência	não entendeu
O1	1307	polvo		1	palavra	não entendeu
O1	1055	alfândega		1	palavra	não entendeu
O1	463	blindado	blindado	2	palavra	acertou
O1	1162	zero		1	palavra	não entendeu
O1	592	manhã linda	manhã linda	2	sequência	acertou
O1	1123	cansado	cansado	3	palavra	acertou
O1	711	atum jeitoso		1	sequência improvável	não entendeu
O1	1524	ramo	rialto	2	palavra	não entendeu
O1	835	apneia	apneia	3	palavra	acertou
O1	1407	novo	novo	3	palavra	acertou
O1	276	pó nhónhó	pó nhonhó	2	sequência improvável	acertou
O1	1566	ocidente	ocidente	3	palavra	acertou
O1	403	situar	situar	3	palavra	acertou
O1	1400	menino	menino	3	palavra	acertou
O1	318	corro	corro	2	palavra	acertou
O1	1057	afundar	afundar	3	palavra	acertou
O1	454	sim antigo	sim antigo	2	sequência improvável	acertou
O1	750	põe chave	pó inchado	2	sequência improvável	sonoridade próxima
O1	103	enciclopédia	enciclopédia	2	palavra	acertou
O1	1580	pato	bato/rato	2	palavra	troca de oclusivas
O1	931	pedinte	presidente	2	palavra	não entendeu
O1	692	atum entre		1	sequência improvável	não entendeu
O1	361	pai esse		1	sequência improvável	não entendeu
O1	927	doce	doce	2	palavra	acertou
O1	1215	chão	chão	3	palavra	acertou
O1	847	tambor	amor	2	palavra	sonoridade próxima
O1	334	como antiga	como antigo	2	sequência	troca de vogais
O1	607	manhã outono	manhã outono	2	sequência	acertou
O1	1252	agenda	agenda	3	palavra	acertou
O1	1088	árvore	árvore	3	palavra	acertou
O1	737	pó untado	pó impado	2	sequência improvável	sonoridade próxima
O1	94	vê antigo	vi antigo	2	sequência	troca de vogais
O1	963	casa	casa	3	palavra	acertou
O1	80	pé ideal	vê areal	1	sequência	sonoridade próxima
O1	664	bronquite	bronquite	3	palavra	acertou

O1	24	auricular	auricular	2	palavra	acertou
O1	1583	duende	doente	2	palavra	troca de oclusivas
O1	1261	cabisbaixo	cabisbaixo	3	palavra	acertou
O1	701	umbigo	umbigo	3	palavra	acertou
O1	792	são lindos	chão nhonhó	2	sequência	não entendeu
O1	342	absoluto	absoluto	3	palavra	acertou
O1	511	enxame	exame	2	palavra	não entendeu
O1	1535	ranço	rancho	2	palavra	troca de fricativas
O1	1161	azedo	azedo	2	palavra	acertou
O1	658	põe	põe	3	palavra	acertou
O1	428	pau só	alço	2	palavra	não entendeu
O1	703	vagabundo	vagabundo	3	palavra	acertou
O1	1492	prenda	duende	2	palavra	não entendeu
O1	1092	vento	vento	2	palavra	acertou
O2	502	assentir	está a sentir	2	palavra	sonoridade próxima
O2	583	escândalo	escândalo	3	palavra	acertou
O2	628	dançarina	dançarina	3	palavra	acertou
O2	191	agitar	a girar	2	palavra	troca oclusiva vibrante
O2	90	céu	cedo	2	palavra	sonoridade próxima
O2	422	autocarro	autocarro	3	palavra	acertou
O2	256	pó ondulado	ondulado	2	sequência improvável	só entendeu a segunda parte
O2	1395	amnistia		1	palavra	não entendeu
O2	380	pai pobre		1	sequência	não entendeu
O2	1032	glicose	glicose	3	palavra	acertou
O2	15	vi antiguidades	vi antiguidades		sequência	acertou
O2	1309	malzinho	mauzinho	2	palavra	semivocalização
O2	687	atum otono		1	sequência improvável	não entendeu
O2	1175	zangar	zangar	3	palavra	acertou
O2	265	criptografia	criptografia	3	palavra	acertou
O2	248	pó uma	home	1	palavra	não entendeu
O2	307	ovino	gago	1	palavra	não entendeu
O2	675	com notícias	com notícias	2	sequência	acertou
O2	1368	locomotiva	locomotiva	3	palavra	acertou
O2	783	são doces	são doces	2	sequência	acertou
O2	1304	polca	polca	3	palavra	acertou
O2	736	põe ontem	põe ontem	2	sequência improvável	acertou
O2	731	põe interesse	põe interesse	2	sequência improvável	acertou
O2	720	põe isso	põe isso	2	sequência improvável	acertou
O2	1441	punheta	punheta	3	palavra	acertou
O2	8	vi urso	vi urso	2	sequência	acertou
O2	475	sim nádia	sim nádia	2	sequência improvável	acertou
O2	78	vê ricas	vê riscas	2	sequência improvável	acertou
O2	1571	indecente	indecente	3	palavra	acertou
O2	883	cantado	antado	2	palavra	elisão do som inicial
O2	639	manhã	manhã	2	palavra	acertou
O2	317	aeroporto	aeroporto	2	palavra	acertou
O2	1172	cinzento	cinzento	3	palavra	acertou
O2	601	manhã esse	manhã isso	2	sequência improvável	troca de vogais

O2	671	com jeito	conceito	2	palavra	troca de fricativas
O2	1205	xenofobia	xenofobia	3	palavra	acertou
O2	588	dançarina	dançarina	3	palavra	acertou
O2	992	ciclo	ciclo	3	palavra	acertou
O2	310	ostra	ostra	3	palavra	acertou
O2	1446	nhónhó	nhónhó	2	palavra	acertou
O2	910	ketchup	ketchup	3	palavra	acertou
O2	1010	egoísmo	egoísmo	3	palavra	acertou
O2	374	comerciante	comerciante	3	palavra	acertou
O2	877	bronquite	bronquite	3	palavra	acertou
O2	1489	ébrio	ébrio	3	palavra	acertou
O2	1048	sinfonia	sinfonia		palavra	acertou
O2	626	manhã fechada	manhã sem enchada	1	sequência improvável	sonoridade próxima
O2	1003	cigarro	cigarro	3	palavra	acertou
O2	1117	palavra	palavra	3	palavra	acertou
O2	635	manhã nova	manhã novo	2	sequência	troca de vogais
O3	985	background		1	palavra	não entendeu
O3	181	acabar	acabar	2	palavra	acertou
O3	1265	rasgar	vais regar	1	palavra	sonoridade próxima
O3	1303	soldado	soldado	3	palavra	acertou
O3	321	como esse	como esse	2	sequência	acertou
O3	436	pau nhónhó	dá-lhe nhónhó	1	sequência improvável	só entendeu a segunda parte
O3	1001	guê	guê	3	palavra	acertou
O3	95	vê antigo	vê antigo	3	sequência	acertou
O3	142	cátia	cátia	3	palavra	acertou
O3	433	viu-lhe	viu-lhe	2	sequência	acertou
O3	654	som antigo	só antigo	2	sequência improvável	desnasalização
O3	764	são aquilo	são aquilo	3	sequência	acertou
O3	954	admite	admite	2	palavra	acertou
O3	1326	palhota	palhota	3	palavra	acertou
O3	1528	carro	carro	3	palavra	acertou
O3	45	vê e	ver	1	palavra	sonoridade próxima
O3	975	alcançar	alcancer	2	palavra	troca de vogais
O3	1121	sê	sê	3	palavra	acertou
O3	147	cravo	cravo	2	palavra	acertou
O3	679	som antigo	som	3	sequência improvável	só entendeu o som inicial
O3	600	manhã imensa	manhã imensa	2	sequência	acertou
O3	265	criptografia	criptografia	3	palavra	acertou
O3	722	põe égua	põe égua	2	sequência	acertou
O3	1529	rialto	rialto	3	palavra	acertou
O3	231	de jaguar	de nada	1	sequência improvável	não entendeu
O3	710	atum chave	atum chave	2	sequência improvável	acertou
O3	912	atletismo	ciclo	1	palavra	não entendeu
O3	1521	rêgo	rêgo	2	palavra	acertou
O3	1023	amigdalite	amigdalite	2	palavra	acertou
O3	395	pai novo	dai novo	2	sequência	troca de oclusivas
O3	563	manhã água	manhã água	2	sequência improvável	acertou

O3	300	sopa	sopa	2	palavra	acertou
O3	784	são coisas	são coisas	2	sequência	acertou
O3	741	põe baixo	põe bege	1	sequência	não entendeu
O3	976	contrato	contrato	3	palavra	acertou
O3	113	velhota	velhota	2	palavra	acertou
O3	502	assentir	da sentir	2	palavra	sonoridade próxima
O3	972	quente	quente	2	palavra	acertou
O3	279	pó	ó	2	palavra	elisão do som inicial
O3	393	foi-lhe	foi-lhe	2	sequência	acertou
O3	809	piada	piada	3	palavra	acertou
O3	743	põe disso	põe disso	2	sequência	acertou
						acrescento de som inicial
O3	1506	arfar	farfar	1	palavra	
O3	642	som égua	som égua	2	sequência improvável	acertou
O3	658	põe	põe	2	palavra	acertou
O3	457	sim untado	cintado	1	sequência improvável	sonoridade próxima
		manhã				
O3	630	chuvosa	manhã chuvosa	2	sequência	acertou
O3	804	aparelho	aparelho	2	palavra	acertou
O3	708	atum seco	atum seco	2	sequência improvável	acertou
O3	746	põe feliz	põe feliz	2	sequência improvável	acertou
O4	595	manhã nova	manhã nova	2	sequência	acertou
O4	117	aberto	aberto	2	palavra	acertou
O4	319	avô		1		não entendeu
O4	79	sê	sid	1	palavra	não entendeu
O4	194	claramente	claramente	2	palavra	acertou
O4	187	salada viva	selada viva	2	sequência improvável	troca de vogais
O4	599	manhã	manha	2	palavra	desnasalização
O4	74	vê-me	vê-me	2	palavra	acertou
O4	461	simbiose	simbiose	2	palavra	acertou
O4	162	aeroporto	aeroporto	3	palavra	acertou
O4	104	biblioteca	biblioteca	3	palavra	acertou
O4	374	comerciante	comerciante	3	palavra	acertou
O4	1492	prenda	prenda	3	palavra	acertou
O4	100	pé positivo	é positivo	2	sequência improvável	elisão do som inicial
O4	422	autocarro	autocarro	2	palavra	acertou
O4	1598	rato	rato	3	palavra	acertou
O4	796	são nhónhó	são nhónhó	3	sequência improvável	acertou
O4	464	zinco	vinco	2	palavra	troca de fricativas
O4	1531	rindo	rindo	2	palavra	acertou
O4	209	disse iate	disse diate	2	sequência improvável	sonoridade próxima
O4	152	maldade	mal deve	1	palavra	sonoridade próxima
O4	23	cândido	andido	2	palavra	elisão do som inicial
O4	460	simples	simples	2	palavra	acertou
O4	1305	salgado	delgado	2	palavra	sonoridade próxima
O4	1575	ambição	ambição	3	palavra	acertou
O4	27	invisível	invisível	3	palavra	acertou
O4	1310	mal chave	mal chave	2	sequência improvável	acertou
O4	895	tanto	anto	2	palavra	elisão do som inicial
O4	390	queixo	queixo	2	palavra	acertou

O4	1314	malmequer	malmequer	2	palavra	acertou
O4	420	pau pequeno	albequeno	1	sequência	só entendeu o fim
O4	1205	xenofobia	xenofobia	3	palavra	acertou
O4	1054	alfândega		1		não entendeu
O4	342	absoluto	absoluto	2	palavra	acertou
O4	962	quero	quero	3	palavra	acertou
O4	967	côto	côto	3	palavra	acertou
O4	1017	segundo	segundo	3	palavra	acertou
O4	272	golpe	golde	2	palavra	troca de oclusivas
O4	862	obter	obter	3	palavra	acertou
O4	1096	vontade	vão ter	2	palavra	não entendeu
O4	1083	novato	novato	2	palavra	acertou
O4	1402	neto	neto	2	palavra	acertou
O4	20	pipoca	pipoca	3	palavra	acertou
O4	1245	vegetal	vegetal	3	palavra	acertou
O4	1481	careta	cadete	2	palavra	não entendeu
O4	1085	suave	suave	3	palavra	acertou
O4	118	guerra	guerra	3	palavra	acertou
O4	113	velhota	velhote	2	palavra	troca de vogais
O4	363	piada	piada	3	palavra	acertou
O4	732	põe entre	pó entre	2	sequência improvável	desnasalização
O5	1056	afonso	arfar	2	palavra	não entendeu
O5	753	põe-lhe	põe-lhe	2	palavra	acertou
O5	238	de Rita	visita	2	sequência improvável	só entendeu o fim
O5	217	de untada	adiantada	2	sequência improvável	sonoridade próxima
O5	851	bimbo	bimbo	2	palavra	acertou
O5	1160	sozinho	sozinho	2	palavra	acertou
O5	456	sim ondulado	simulado	1	sequência improvável	sonoridade próxima
O5	1001	guê	guê	2	palavra	acertou
O5	385	pai gago	vai olhando	1	sequência improvável	não entendeu
O5	1563	activo	activo	3	palavra	acertou
O5	367	miolo	miolo	3	palavra	acertou
O5	1064	kafka	casca	2	palavra	troca de fricativas
O5	215	de antiga	de antigo	3	sequência improvável	troca de vogais
O5	35	clinico	clinico	3	palavra	acertou
O5	186	gafanhoto	gafanhoto	3	palavra	acertou
O5	110	flecha	flecha	2	palavra	acertou
O5	1363	mala	mala	2	palavra	acertou
O5	288	avô único	aro único	1	sequência	troca fricativa vibrante
O5	51	vê interesse	vê interesse	2	sequência	acertou
O5	857	vagabundo	vagabundo	3	palavra	acertou
O5	441	sim esse	chinês	2	sequência	não entendeu
O5	589	canzarrão	canzarrão	3	palavra	acertou
O5	20	pipoca	pipoca	3	palavra	acertou
O5	406	quota	flauta	2	palavra	não entendeu
O5	910	ketchup	ketchup	3	palavra	acertou
O5	254	pó antigo	pó antigo	2	sequência improvável	acertou
O5	950	stands	distantes	2	palavra	sonoridade próxima
O5	1567	oito	oito	2	palavra	acertou

O5	243	pó água	palavra	1	sequência improvável	sonoridade próxima
O5	306	avô feliz	avô feliz	2	sequência	acertou
O5	607	manhã outono	manhã outono	3	sequência improvável	acertou
O5	706	atum feliz	atum feliz	3	sequência improvável	acertou
O5	772	são entre	são entre	2	sequência improvável	acertou
O5	972	quente	quente	3	palavra	acertou
O5	281	avô esse		1	sequência	não entendeu
O5	3	bloquear	bloquear	2	palavra	acertou
O5	1263	desde	desde	2	palavra	acertou
O5	340	lupa	lupa	2	palavra	acertou
O5	251	pó inteiro	pau inteiro	2	sequência improvável	ditongação
O5	651	som inteiro	só inteiro	2	sequência improvável	desnasalização
O5	652	som entre	só entre	2	sequência improvável	desnasalização
O5	84	pé aguado		1	sequência improvável	não entendeu
O5	1505	cargo	cargo	2	palavra	acertou
O5	239	de	be	2	palavra	troca de oclusivas
O5	614	manhã antiga	manhã antiga	3	sequência improvável	acertou
O5	269	tuberculose	tuberculose	3	palavra	acertou
O5	657	som untado	sentado	2	sequência improvável	sonoridade próxima
O5	639	manhã	manha	2	palavra	desnasalização
O5	808	camponês	camponês	2	palavra	acertou
O5	270	carochinha	carochinha	2	palavra	acertou

Anexo 3

Tabela de resultados voz com pronúncia da variante do Porto

ouvinte	Nº	estímulo	o que ouviram	classificação	tipo	tipo de erro
	793	são-lhe	são-lhe	2	sequência	acertou queda do som [l] em posição medial
OP1	1307	polvo	povo	1	palavra	acertou
OP1	1055	alfândega	alfândega	3	palavra	acertou
OP1	463	blindado	blindado	3	palavra	acertou
OP1	1162	zero	zero	3	palavra	acertou
OP1	592	manhã linda	manhã linda	2	sequência	acertou
OP1	1123	cansado	cansado	2	palavra	acertou
OP1	711	atum jeitoso	atum jeitoso	2	sequência improvável	acertou
OP1	1524	ramo	rialto	1	palavra	não entendeu
OP1	835	apneia	apneia	2	palavra	acertou
OP1	1407	novo	novo	3	palavra	acertou troca de oclusiva
OP1	276	pó nhónhó	ronhonhó	2	sequência improvável	vibrante
OP1	1566	ocidente	ocidente	3	palavra	acertou
OP1	403	situar	situar	3	palavra	acertou
OP1	1400	menino	menino	2	palavra	acertou
OP1	318	corro	corro	2	palavra	acertou
OP1	1057	afundar	afundar	2	palavra	acertou
OP1	454	sim antigo	sim antigo	2	sequência improvável	acertou
OP1	750	põe chave	põe chave	2	sequência improvável	acertou
OP1	103	enciclopédia	enciclopédia	2	palavra	acertou
OP1	1580	pato	pato	3	palavra	acertou
OP1	931	pedinte	pedinte	3	palavra	acertou
OP1	692	atum entre	atum entre	3	sequência improvável	acertou
OP1	361	pai esse	pai esse	2	sequência improvável	acertou
OP1	927	doce	doce	2	palavra	acertou
OP1	1215	chão	chão	2	palavra	acertou
OP1	847	tambor	tambor	3	palavra	acertou
OP1	334	como antiga	como antiga	2	sequência improvável	acertou
OP1	607	outono	manhã outono	2	sequência	acertou
OP1	1252	agenda	agenda	2	palavra	acertou
OP1	1088	árvore	árvore	3	palavra	acertou só entendeu a sequência final
OP1	737	pó untado	punho untado	2	sequência improvável	acertou
OP1	94	vê antigo	vê antigo	2	sequência improvável	acertou
OP1	963	casa	casa	2	palavra	acertou
OP1	80	pé ideal	pé ideal	2	sequência	acertou
OP1	664	bronquite	bronquite	2	palavra	acertou
OP1	24	auricular	auricular	2	palavra	acertou
OP1	1583	duende	duende	2	palavra	acertou

OP1	1261	cabisbaixo	cabisbaixo	2	palavra	acertou
OP1	701	umbigo	umbigo	3	palavra	acertou
OP1	792	são lindos	são.....	1	sequência	só entendeu a sequência inicial
OP1	342	absoluto	absoluto beija-me/enxame/	2	palavra	acertou
OP1	511	enxame	exame	1	palavra	não entendeu
OP1	1535	ranço	raso	1	palavra	não entendeu
OP1	1161	azedo	azedo	2	palavra	acertou
OP1	658	põe	põe	2	palavra	acertou
OP1	428	pau só	pau zó	1	palavra	troca de fricativas
OP1	703	vagabundo	vagabundo	2	palavra	acertou
OP1	1492	prenda	prende	1	palavra	troca de vogais
OP1	1092	vento	vento	3	palavra	acertou
OP2	502	assentir		1	palavra	não entendeu
OP2	583	escândalo		1	palavra	não entendeu
OP2	628	dançarina	dançarina	2	palavra	acertou
OP2	191	agitar	agitar	2	palavra	acertou
OP2	90	céu	cedo	1	palavra	não entendeu
OP2	422	autocarro	autocarro	2	palavra	acertou
OP2	256	pó ondulado		1	sequência improvável	não entendeu
OP2	1395	amnístia	amnístia	2	palavra	acertou
OP2	380	pai pobre	pai pobre	2	sequência	acertou
OP2	1032	glicose	glicose	3	palavra	acertou
OP2	15	antiguidades		1	sequência improvável	não entendeu
OP2	1309	malzinho	mauzinho	1	palavra	semivocalização
OP2	687	atum outono	atum outono	2	sequência improvável	acertou
OP2	1175	zangar	zangar	2	palavra	acertou
OP2	265	criptografia	criptografia	2	palavra	acertou
OP2	248	pó uma	pó uma	2	palavra	acertou
OP2	307	ovino	ovino	3	palavra	acertou
OP2	675	com notícias	com notícias	2	sequência	acertou
OP2	1368	locomotiva	locomotiva	3	palavra	acertou
OP2	783	são doces	são doces	2	sequência	acertou
OP2	1304	polca	polca	2	palavra	acertou
OP2	736	põe ontem	põe ontem	2	sequência improvável	acertou
OP2	731	interesse	põe interesse	2	sequência improvável	acertou
OP2	720	põe isso	põe isso	2	sequência	acertou
OP2	1441	punheta	punheta	3	palavra	acertou
OP2	8	vi urso	vi isso	1	sequência improvável	não entendeu
OP2	475	sim nádia	sim nádia	2	sequência improvável	acertou
OP2	78	vê riscas	vê riscas	2	palavra	acertou
OP2	1571	indecente	indecente	2	palavra	acertou
OP2	883	cantado	cantado	3	palavra	acertou
OP2	639	manhã	manhã	3	palavra	acertou
OP2	317	aeroporto	aeroporto	3	palavra	acertou
OP2	1172	cinzento	cinzento	3	palavra	acertou
OP2	601	manhã esse	manhã esse	2	sequência	acertou
OP2	671	com jeito	com jeito	2	palavra	acertou
OP2	1205	xenofobia	xenofobia	2	palavra	acertou
OP2	588	dançarina	dançarina	3	palavra	acertou

OP2	992	ciclo	ciclo	3	palavra	acertou
OP2	310	ostra	ostra	3	palavra	acertou
OP2	1446	nhónhó	nhónhó	2	palavra	acertou
OP2	910	ketchup		1	palavra	não entendeu
OP2	1010	egoísmo	egoísmo	2	palavra	acertou
OP2	374	comerciante	comerciante	2	palavra	acertou
OP2	877	bronquite	bronquite	2	palavra	acertou
OP2	1489	ébrio	ébrio	2	palavra	acertou
OP2	1048	sinfonia	sinfonia	2	palavra	acertou
OP2	626	fechada	manhã fechada	2	sequência improvável	acertou
OP2	1003	cigarro	cigarro	3	palavra	acertou
OP2	1117	palavra	palavra	3	palavra	acertou
OP2	635	manhã nova	manhã nova	2	sequência	acertou
OP3	985	background	background	2	palavra	acertou
OP3	181	acabar	acabar	3	palavra	acertou
OP3	1265	rasgar	rasgar	3	palavra	acertou
OP3	1303	soldado	soldado	3	palavra	acertou
OP3	321	como esse	como...	1	sequência improvável	só entendeu a sequência inicial
OP3	436	pau nhónhó	au nhónhó	1	sequência improvável	elisão do som inicial
OP3	1001	guê	guê	2	palavra	acertou
OP3	95	vê antigo	vê antigo	2	sequência	acertou
OP3	142	cátia	cátia	2	palavra	acertou
OP3	433	viu-lhe	vê-lhe	1	sequência	não entendeu
OP3	654	som antigo	sol antigo	1	sequência improvável	desnasalização
OP3	764	são aquilo	são aquilo	2	sequência	acertou
OP3	954	admite	admite	2	palavra	acertou
OP3	1326	palhota	palhota	2	palavra	acertou
OP3	1528	carro	carro	3	palavra	acertou
OP3	45	vê e	vêe	2	palavra	acertou
OP3	975	alcançar	alcançar	3	palavra	acertou
OP3	1121	sê	sê	2	palavra	acertou
OP3	147	cravo	cravo	3	palavra	acertou
OP3	679	som antigo	som antigo	2	palavra	acertou
OP3	600	imensa	manhã imensa	2	sequência improvável	acertou
OP3	265	criptografia	criptografia	2	palavra	acertou
OP3	722	põe égua	põe água	1	sequência improvável	troca de vogais
OP3	1529	rialto	rialto	3	palavra	acertou
OP3	231	de jaguar	de nada	1	sequência improvável	não entendeu
OP3	710	atum chave	atum chave	2	sequência improvável	acertou
OP3	912	atletismo	atletismo	2	palavra	acertou
OP3	1521	rêgo	rêgo	3	palavra	acertou
OP3	1023	amigdalite	amigdalite	2	palavra	acertou
OP3	395	pai novo	pai novo	2	sequência	acertou
OP3	563	manhã água	manhã água	2	sequência improvável	acertou
OP3	300	sopa	sopa	3	palavra	acertou
OP3	784	são coisas	são coisas	2	sequência	acertou
OP3	741	põe baixo	põe baixo	2	sequência improvável	acertou
OP3	976	contrato	contrato	3	palavra	acertou
OP3	113	velhota	velhota	2	palavra	acertou
OP3	502	assentir	assentir	2	palavra	acertou

OP3	972	quente	quente	2	palavra	acertou
OP3	279	pó	pó	3	palavra	acertou
OP3	393	foi-lhelhe	2	sequência improvável	só entendeu a sequência final
OP3	809	piada	piada	3	palavra	acertou
OP3	743	põe disso	põe disso	2	sequência	acertou
OP3	1506	arfarfar	1	palavra	só entendeu a sequência final
OP3	642	som égua	som égua	2	sequência improvável	acertou
OP3	658	põe	põe	2	palavra	acertou
OP3	457	sim untado	sim untado	2	sequência improvável	acertou
		manhã				
OP3	630	chuvosa	manhã chuvosa	2	sequência	acertou
OP3	804	aparelho	aparelho	3	palavra	acertou
OP3	708	atum seco	atum seco	2	sequência improvável	acertou
						troca oclusiva
OP3	746	põe feliz	põe veliz	1	sequência improvável	fricativa
OP4	595	manhã nova	manhã nova	2	sequência improvável	acertou
OP4	117	aberto	inverto	1	palavra	não entendeu
OP4	319	avô	avô	3	palavra	acertou
OP4	79	sê	sê	3	palavra	acertou
OP4	194	claramente	claramente	2	palavra	acertou
OP4	187	salada viva	selada viva	1	sequência improvável	troca de vogais
OP4	599	manhã	manhã	2	palavra	acertou
OP4	74	vê-me	vê-me	2	palavra	acertou
OP4	461	simbiose	simbiose	3	palavra	acertou
OP4	162	aeroporto	aeroporto	3	palavra	acertou
OP4	104	biblioteca	biblioteca	3	palavra	acertou
OP4	374	comerciante	comerciante	3	palavra	acertou
OP4	1492	prenda	prenda	3	palavra	acertou
OP4	100	pé positivo	pé positivo	2	sequência improvável	acertou
OP4	422	autocarro	autocarro	3	palavra	acertou
OP4	1598	rato	rato	3	palavra	acertou
OP4	796	são nhónhó	são nhónhó	2	sequência improvável	acertou
OP4	464	zinco	zinco	2	palavra	acertou
OP4	1531	rindo	rindo	2	palavra	acertou
OP4	209	disse iate	disse iate	2	sequência improvável	acertou
OP4	152	maldade	maldade	2	palavra	acertou
OP4	23	cândido	cândido	2	palavra	acertou
OP4	460	simples	simples	2	palavra	acertou
OP4	1305	salgado	salgado	2	palavra	acertou
OP4	1575	ambição	ambição	2	palavra	acertou
OP4	27	invisível	invisível	2	palavra	acertou
OP4	1310	mal chave	mal chave	2	sequência improvável	acertou
OP4	895	tanto	tanto	2	palavra	acertou
OP4	390	queixo	queijo	1	palavra	troca de fricativas
OP4	1314	malmequer	malmequer	2	palavra	acertou
		pau				
OP4	420	pequeno	pau pequeno	3	sequência	acertou
OP4	1205	xenofobia	xenofobia	2	palavra	acertou
OP4	1054	alfândega	alfândega	3	palavra	acertou
OP4	342	absoluto	absoluto	3	palavra	acertou
OP4	962	quero	quero	2	palavra	acertou

OP4	967	côto	côto	3	palavra	acertou
OP4	1017	segundo	segundo	2	palavra	acertou
OP4	272	golpe	golpe	2	palavra	acertou
OP4	862	obter	obter	2	palavra	acertou
OP4	1096	vontade	vontade	2	palavra	acertou
OP4	1083	novato	novato	2	palavra	acertou
OP4	1402	neto	neto	2	palavra	acertou
OP4	20	pipoca	pipoca	3	palavra	acertou
OP4	1245	vegetal	vegetal	2	palavra	acertou
OP4	1481	careta	careta	2	palavra	acertou
OP4	1085	suave	suave	2	palavra	acertou
OP4	118	guerra	guerra	2	palavra	acertou
OP4	113	velhota	velhote	1	palavra	troca de vogais
OP4	363	piada	piada	2	palavra	acertou
OP4	732	põe entre	põe entre	2	sequência improvável	acertou
OP5	1056	afonso		1	palavra	não entendeu
OP5	753	põe-lhe	ponha	2	palavra	não entendeu
OP5	238	de Rita		1	sequência improvável	não entendeu
OP5	217	de untada	de untada	2	sequência improvável	acertou
OP5	851	bimbo	bimbo	3	palavra	acertou
OP5	1160	sozinho	sozinho	3	palavra	acertou
		sim				
OP5	456	ondulado	sim ondulado	2	sequência improvável	acertou
OP5	1001	guê	guê	2	palavra	acertou
OP5	385	pai gago	pai gago	2	sequência	acertou
OP5	1563	activo	activo	2	palavra	acertou
OP5	367	miolo	miolo	3	palavra	acertou
OP5	1064	kafka	kafka	3	palavra	acertou
OP5	215	de antiga	de antiga	3	sequência improvável	acertou
OP5	35	clínico	clínico	3	palavra	acertou
OP5	186	gafanhoto	gafanhoto	2	palavra	acertou
OP5	110	flecha	flecha	2	palavra	acertou
OP5	1363	mala	mala	3	palavra	acertou
						só entendeu a
OP5	288	avô únicoúnico	1	sequência	sequência final
OP5	51	vê interesse	vê interesse	2	sequência	acertou
OP5	857	vagabundo	vagabundo	2	palavra	acertou
OP5	441	sim esse	sim esse	2	sequência	acertou
OP5	589	canzarrão	canzarrão	2	palavra	acertou
OP5	20	pipoca	pipoca	3	palavra	acertou
OP5	406	quota	quota	2	palavra	acertou
OP5	910	ketchup	ketchup	3	palavra	acertou
OP5	254	pó antigo	pau antigo	1	sequência	ditongação
OP5	950	stands	estante	1	palavra	sonoridade próxima
OP5	1567	oito	oito	3	palavra	acertou
OP5	243	pó água	pau água	1	sequência improvável	ditongação
OP5	306	avô feliz	avô feliz	2	sequência	acertou
		manhã				
OP5	607	outono	manhã outono	2	sequência improvável	acertou
OP5	706	atum feliz	atum feliz	2	sequência improvável	acertou
OP5	772	são entre		1	sequência improvável	não entendeu
OP5	972	quente	quente	3	palavra	acertou
OP5	281	avô esse		1	sequência improvável	não entendeu

OP5	3	bloquear	bloquear	3	palavra	acertou
OP5	1263	desde	desde	2	palavra	acertou
OP5	340	lupa	lupa	3	palavra	acertou
OP5	251	pó inteiro	pau inteiro	1	sequência improvável	ditongação
OP5	651	som inteiro	som inteiro	2	sequência improvável	acertou
OP5	652	som entre	som entre	2	sequência improvável	acertou
OP5	84	pé aguado	pé aguado	2	sequência improvável	acertou
OP5	1505	cargo	cargo	2	palavra	acertou
OP5	239	de manhã		1	palavra	não entendeu
OP5	614	antiga	manhã antiga	2	sequência improvável	acertou
OP5	269	tuberculose	tuberculose	2	palavra	acertou
OP5	657	som untado	som untado	2	palavra	acertou
OP5	639	manhã	manhã	3	palavra	acertou
OP5	808	camponês	camponês	3	palavra	acertou
OP5	270	carochinha	carochinha	3	palavra	acertou

Anexo 4 - frases tratadas para o teste 6.2

Standard

frase 1

#lidEr p6l@StinianuSILd@z6p6rEs auS s@te~ t6isi~ ku 6nuS
#

frase 2

#p6tro~j~ S 6kuz6~w~ p6rtiduZ d6tuar puri~ t@res @ prOpriu
#

frase 3

#kwaz@m@tad@d6S f6mili6S portug ez6S# 6bit6 6~j~ kaz6
S fri6S#

frase 4

#kum@so u Zulg6me~ tu#du prusEs u majZm@diatiku duz u
ltimuz 6nuS#

frase 5

#fut@bOl klub@ duportu foj oZ 6gr6siadupelu guvernu#ko~
6 m@daL 6 d@ mEritu turiStiku#

frase 6

#eSt E u~SpE takulud@s6p6tiadu#po ku o nad6 ko~ ve~ siun
al#

frase 7

#susiEdad@di~ furm6s6~w~#E fu~ d6me~ tal p6r6 ko~ b6ter m
uruzidad@#

frase 8

#St6muS ko~ 6 s@lE s6~w~n6siunal#6~j~ toduz uZ d@z6fiwS
#

Frases 9

#liZbo 6~ tig6vibro ko~ u d@Sfil@ k6rn6v6leSku#

frase 10

#bLOku diS k 6igr6Z6#n6~w~ dEv@tumar puziso~j~ Spulitik
6S#

frase 11

#kami6~w~d@Sko~truladu#mat6tri~t6iojtu p@s o6S#6~j~
6~gO16#

frase 12

#6m@nin6duZ b6jZuS#

frase 13

#u d@ver d@vutar#

Porto

frase 1

#lidEr@p6l@StinianuSILd@z6p6rEs auS s@te~t6isi~ku an
uS#

frase 2

#p6tro~j~S6kuz6~w~p6rtiduZ d6tuar@pur i~t@res @prOpri
u#

frase 3

#kwaz@m@tad@d6S f6mili6S portug ez6S#6bit66~j~J@ka
z6S fri6S#

frase 4

#kum@sow u Zulg6me~tu#du prusEs u majZm@diatiku duz
ultimuz anuS#

frase 5

#fut@bOl@k@lub@dupw6rtu foj @ojZ 6gr6siadupelu gub
ernu#ko~6m@dajL 6d@mEritu turiStiku#

frase 6

#eStE u~SpE takulu d@s6p6tiadu#pow ku ow nad6ko~be~s
iunal#

frase 7

#susiEdad@di~furm6s6~w~#Efu~d6me~tal@p6r6ko~b6ter
@muruzidad@#

frase 8

#iStamuS ko~6s@lE s6~w~n6siunal@#6~j~tw6duz uZ d@z6
fiuS#

frase 9

#liZbow6 6~ tig6 bibrow@ ko~ u d@ Sfil@ k6rn6b6leSku#

frase 10

#blOku diS k i6grejZ6 n6~w~ dEb@ tumar@ puziso~j~S pulit
ik6S#

frase 11

#kami6~w~ @ d@ Sko~ truladu#mat6 tri~ t6 i Ojtu p@s ow6S#
6~j~J@ 6~ gOl6#

frase 12

#6m@ nin6 duZ bejZuS#

frase 13

#u d@ ber@ d@ butar@ #